

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 19:48:01 ; Search time 41.6 Seconds
(without alignments)
2323.495 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcaataaaga.....catcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.n2p.model -DEV=x1p
-Q=/abs/ABSSWEB.spool/US09360685/runat_06032006_121316_15362/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOFCI=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p
-USER=US09360685@CCN_1_1_466@runat_06032006_121316_15362 -NCPU=6 -ICPU=3
-NO_MMAP -NB2_SCORES=0 -WAIT -DSRLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	1	CGA1_HELPY
2	1127	96.2	1147	1	P80200 helicobacte
3	1029	87.9	1181	2	Q6VROS_HELPY
4	1029	87.9	1181	2	O6VRW7 helicobacte
5	1029	87.9	1230	2	O4PLIS_HELPY
6	1025.5	87.6	1267	2	O8RNU0_HELPY
7	1016	86.8	1142	2	O52GZ6_HELPY
8	1009	86.2	1180	2	O8KZB3_HELPY
9	1004	85.7	1179	2	O7X4J1_HELPY
10	1003.5	85.7	1179	2	O9L5X9_HELPY
11	992	84.7	1186	2	O5GZS5_HELPY
12	990.5	84.6	1179	2	O9F222_HELPY
13	984	84.0	1182	2	O6VRG7_HELPY
14	982.5	83.9	1394	2	O8RRY0_HELPY
15	972.5	83.0	1327	2	O8KZB0_HELPY
16	972.5	83.0	1148	2	O60FP6_HELPY

17	963	82.2	441	2	O9L7K3_HELPY	O9L7K3 helicobacte
18	962.5	82.2	359	2	O8KZB8_HELPY	O8KZB8 helicobacte
19	962.5	82.2	1183	2	O60FQ5_HELPY	O60FQ5 helicobacte
20	960	82.0	1183	2	O60604_HELPY	O60604 helicobacte
21	959.5	81.9	1186	1	CAGA_HELPY	P55980 helicobacte
22	958.5	81.9	1247	2	O9F223_HELPY	O9F223 helicobacte
23	958.5	81.9	1247	2	O9RFL5_HELPY	O9RFL5 helicobacte
24	955.5	81.6	1156	2	O5D6R2_HELPY	O5D6R2 helicobacte
25	952	81.3	1216	2	O8KZB9_HELPY	O8KZB9 helicobacte
26	948.5	81.0	1247	2	O07910_HELPY	O07910 helicobacte
27	944.5	80.7	408	2	O9L7K4_HELPY	O9L7K4 helicobacte
28	942.5	80.5	1184	2	O60FQ2_HELPY	O60FQ2 helicobacte
29	941	80.4	1222	2	O8KZB8_HELPY	O8KZB8 helicobacte
30	939.5	80.2	359	2	O8KZB8_HELPY	O8KZB8 helicobacte
31	939.5	80.2	1183	2	O60FP5_HELPY	O60FP5 helicobacte
32	939.5	80.2	1183	2	O60FP3_HELPY	O60FP3 helicobacte
33	939.5	80.2	1183	2	O60FP3_HELPY	O60FP3 helicobacte
34	939.5	80.2	1188	2	O60FN9_HELPY	O60FN9 helicobacte
35	938.5	80.1	1190	2	O5D6R1_HELPY	O5D6R1 helicobacte
36	937.5	80.1	1183	2	O60FP9_HELPY	O60FP9 helicobacte
37	934	79.8	391	2	O8KZC1_HELPY	O8KZC1 helicobacte
38	934	79.8	1214	2	O60FQ8_HELPY	O60FQ8 helicobacte
39	930.5	79.5	362	2	O8KZB5_HELPY	O8KZB5 helicobacte
40	930.5	79.5	1183	2	O60FQ4_HELPY	O60FQ4 helicobacte
41	930.5	79.5	1189	2	O60FP4_HELPY	O60FP4 helicobacte
42	926.5	79.1	1320	2	O6VRU6_HELPY	O6VRU6 helicobacte
43	918	78.4	1184	2	O60FP2_HELPY	O60FP2 helicobacte
44	912.5	77.9	445	2	O9L7K5_HELPY	O9L7K5 helicobacte
45	902.5	77.1	1182	1	CGA2_HELPY	P55746 helicobacte

ALIGNMENTS

RESULT 1

CGA1_HELPY STANDARD; PRT: 1147 AA.
ID CGA1_HELPY
AC P80200.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN Name=cga1; Synonyms=cag1;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CTUG 1874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petraccia R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N., Rappunli R.,
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal ulcer";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CTUG 1874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108; DOI=10.1073/pnas.93.25.14648;
RA Censini S., Lange C., Xiang Z., Cradocke J., Ghara P., Borodovsky M.,
RA Rappunli R., Covacci A.;
RT "cag", a pathogenicity island of Helicobacter pylori, encodes type I -
RT specific and disease-associated virulence factors";
RN Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RP [3]
RP PROTEIN SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506;
RP 661-677; 900-914 AND 1062-1077.
RA Herrmann V., Herrmann J., Kist M.;
RL Submitted (APR-1993) to Swiss-Prot.
CC -!- FUNCTION: May be necessary for the transcription, folding, export,
CC or function of the cytotoxin.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration

QY 301 GGTAAATGGATTATCTCAAGCAGACCAACTCTTTCTPAAAACTTTTCGACATCAAG 360
 DB 850 GYAASGlyLeuSerGlnAlaGlnAlaThrThreLeuSerLysAsnPheserApIleLys 869
 QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATACATTAACATTAATGACTCAAAAC 420
 DB 870 LysGlnLeuAsnAlaLysLeuGlyAsnPhesAsnAsnAsnAsnAsnAsnGlnGlyLeuLysAsn 889
 QY 421 GAACCATTTATGCTAAAGTATATATAAAGAAAGCGGGCAGACGACTGACCTTGAGAA 480
 DB 890 GluProIleTyrAlaLysValaLysValaLysValaGlyGlnAlaLaseLeuGluGlu 909
 QY 481 CCCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAAATGACCGCATCAATCAATA 540
 DB 910 ProIleTyrAlaGlnAlaAlaLysValaLysValaLysLysLeuAsnGlnIle 929
 QY 541 GCAAGTGTGGTGGTGTGGTGTGGGCAAGCAGCGGCTTCCCTTGAAGGCAATGATAA 600
 DB 930 AlaSerGlyLeuGlyValaLysGlnAlaLysGlnAlaGlyPheProLeuLysArgHisAspLys 949
 QY 601 GTTGAATCTCAGTAAGGAGGCTTTCAAGAAATCAAGATTTGCTCAAAAATTGAC 660
 DB 950 ValAspAspLeuSerLysValaGlyLeuSerArgAsnGlnGluLeuAlaGlnLysIleAsp 969
 QY 661 AATCTCAATCAAGCGGTATCAGAA 684
 DB 970 AsnLeuAsnGlnAlaValaSerGlu 977

RESULT 3

Q6VR05_HELPLY PRELIMINARY: PRT: 1198 AA.

ID 06VR05_HELPLY PRELIMINARY: PRT: 1198 AA.
 AC 06VR05_7
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Caga.
 GN Name=Caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;
 RN
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Cas2;
 RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
 RA Blomestegren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
 RT "Comparative analysis of the complete cag pathogenicity island
 sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004)
 DR EMBL: AY330637; AAR03881.1; -; Genomic_DNA.
 DR GO: GO:0019534; P:toxin transporter activity; IBA.
 DR InterPro: IPR005169; Caga.
 DR InterPro: IPR004355; IVSec_caga.
 DR Pfam: PF03507; Caga; 1.
 DR PRINTS: PRO1553; TYPR4SSCAGA.
 DR SEQUENCE 1198 AA; 133091 MW; DBBFSB554663BCOE CRC64;

Alignment Scores:

Pred. No.: 6.22e-66 Length: 1198
 Score: 1035.00 Matches: 214
 Percent Similarity: 83.2% Conservative: 4
 Best Local Similarity: 81.7% Mismatches: 10
 Query Match: 88.4% Indels: 34
 DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x Q6VR05_HELPLY (1-1198)

QY 1 AAAAAATGCAAAATPAGGATTTCAGCAAGTAACGCAAGCAAAAAAGCGACTTGAAAAAT 60
 DB 766 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGlnAsn 785
 QY 61 TCCGTTAAAGATGATATCATCAATCAAAAGTAACGATTAAGTTGATTAATCTCAATCA 120

DB 786 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValaAspAsnLeuAsnGln 805
 QY 121 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTTCACTAGCTAGACCAAGCCGAT 180
 DB 806 AlaValSerIleAlaLysAlaThrGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 825
 QY 181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT 240
 DB 826 LeuLysAsnPheserLysGlyGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 845
 QY 241 GCTAGAAAAAATCGAAATATATCAATCCGTTAAGATGAGTGAAATGAACCTTACTC 300
 DB 846 ThrGlyLysAsnSerGlnIleTyrGlnSerValLysAsnGlyValaGlnIleThrLeuVal 865
 QY 301 GGTAAATGGATTATCTCAAGCAGACCAACTCTTTCTPAAAACTTTTCGACATCAAG 360
 DB 866 GYAASGlyLeuSerLysAlaGlnAlaThrThreLeuSerLysAsnPheserApIleLys 885
 QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATACATTAACATTAATGACTCAAAAC 420
 DB 886 LysGlnLeuGlnAlaLysLeuGlyAsnPhesAsnAsnAsnAsnAsnAsnGlnLysLysAsn 905
 QY 421 GAACCATTTATGCTAAAGTATATATAAAGAAAGCGGGCAGACGACTGACCTTGAGAA 480
 DB 906 GluProIleTyrAlaLysValaLysLysLysLysAlaGlyGlnAlaLaseProGlnGlu 925
 QY 481 CCCATTTAGCTCAAGTTGCTAAAGGTAATGCAAAAATGACCGCATCAATCAATA 540
 DB 926 ProIleTyrAlaGlnAlaAlaLysValaLysValaLysLysLeuAsnGlnIle 945
 QY 541 GCAAGTGTGGTGGTGTGGTGTGGGCAAGCAGCGGCTTCCCTTGAAGGCAATGATAA 576
 DB 946 AlaSerGlyLeuGlyValaGlyGlnAlaLysGlnAlaGlyPheProLeuLysArgHisAspLys 965
 QY 576 ----- 576
 DB 966 ValAspAspLeuSerLysValaGlyLeuSerAlaAsnHisGlnProIleTyrAlaThrIle 985
 QY 577 -----TTCCTTTGAAAAAGCATGATAAGTTGATGATCTCAGTAAG 618
 DB 986 AspAspLeuGlyGlySerPheProLeuLysArgHisAspLysValaAspLeuSerLys 1005
 QY 619 GTAGGGCTTTCAAGAAATCAAGAAATTTGCTCAAAAATTGACATCTCAATCAAGCGGTA 678
 DB 1006 ValGlyLeuSerArgAsnGlnGlnLeuAlaGlnLysIleAspAsnLeuAsnGlnAlaVal 1025
 QY 679 TCAGAA 684
 DB 1026 SerGlu 1027

RESULT 4

Q6VR07_HELPLY PRELIMINARY: PRT: 1181 AA.

ID 06VR07_HELPLY PRELIMINARY: PRT: 1181 AA.
 AC 06VR07_7
 DT 05-JUL-2004 (TREMBlrel. 27, Created)
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE Caga.
 GN Name=Caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;
 RN
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Cas73;
 RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
 RA Blomestegren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
 RT "Comparative analysis of the complete cag pathogenicity island
 sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004)
 DR EMBL: AY330639; AAR03909.1; -; Genomic_DNA.

DR GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro:IPR005169; CgA.
 DR InterPro:IPR004355; IYSec_CgA.
 DR Pfam:PF03507; CgA.1
 DR PRINTS:PRO1553; TYPE4SCGA.
 SQ SEQUENCE 1181 AA; 131427 MW; A991CAFD617CDE3E CRC64;

Alignment Scores:
 Pred. No.: 1,68e-65 Length: 1181
 Score: 1029.00 Matches: 213
 Percent Similarity: 82.8% Conservative: 4
 Best Local Similarity: 81.3% Mismatches: 11
 Query Match: 87.9% Indels: 34
 DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x O6VM7_HELPY (1-1181)

```

QY 1 AAAATGCGCAAAATTAAGGATTTTCAGCAGGTAAAGCAAAAGGACCTTGAAT 60
DB LysanGlyLysanLysanLysanPheSerValThrglnAlaLysSerAplLeuLysan 765
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGGTAAGGTAATGATCAATCA 120
DB SerLleYasPheValIleIleAenGlnLysIleThraPlyValAenLysSer 789
QY 121 GCGGTATCATGTGCTAAAGCAAGGATTTTCAGTGGGTAAAGGCTTACCGCAT 180
DB AlaValSerValAlaLysAlaThrgLysAphSerValGlnValLeuAlaGly 809
QY 181 CTCAAAAATTTCTCAAGAGCAATGCGCCCAAGACTCAAAAAATGAATCTCAAT 240
DB LeuLysanPheSerLysGlnLysanLysanLysanLysanLysanLysan 829
QY 241 GCTGAAAAAATCTGAAATATATCAATCGTTAAAGTGGTGAATGAACCTTATGC 300
DB ThrGlyLysLysSerGlnLysIleThraSerValLysanGlyValAenGlyThraLeuAl 849
QY 301 GGTATGCGGTATTCGAAAGCAAGCAACTTTTCTAAAACTTTTGGACATCAG 360
DB GlyAenGlyLysSerGlnLysanLysanLysanLysanLysanLysanLysan 869
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATACAAATTAATGACTCAAAAAC 420
DB LysGlnLysanLysanLysanLysanLysanLysanLysanLysanLysan 889
QY 421 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGA 480
DB GlnProLysLysValLysanLysanLysanLysanLysanLysanLysan 909
QY 481 CCGATTACCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 540
DB ProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 929
QY 541 GCAAGTGGTTGGTGTGTAGGCGCAAGCCGGC----- 576
DB 930 AlaSerGlyLysGlnLysValGlyGlnAlaLysLysPheProLeuLysanLysanLysan 949
QY 576 ----- 576
DB 950 ValAapPheLysSerLysValGlyLysSerAlaAenProGlnProLysLysLysLys 969
QY 577 -----TTCCCTTGAAGGCAATGATTAAGTAAAGTAAAGTAAAGTAAAG 618
DB 970 AapGlnLysanLysLysProPheProLysLysLysLysLysLysLysLysLysLys 989
QY 619 GTAGGCTTTCAAGAAATCAAGAAATGCTCAAGAAATGCAATCTCAATCAAGCGGTA 678
DB 990 ValGlyLysSerLysanGlnLysLysLysLysLysLysLysLysLysLysLysLys 1009
QY 679 TCAGAA 684
DB 1010 SerGln 1011

```

RESULT 5
 OAPL15_HELPY
 ID OAPL15_HELPY PRELIMINARY; PRT: 1230 AA.
 AC OAPL15;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Cytochrome-associated protein A.
 GN Hmnc-cgA.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteriaceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=7bgs;
 RA Kim S.Y., Blaser M.J., Lee Y.C., Pilling M.H.;
 RT "Helicobacter pylori stimulates matrix metalloproteinase-1 secretion
 RT from gastric epithelial cells via CagA-dependent and independent
 RT mechanisms: requirement for Erk activation."
 RU Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D0067454; AAY68033.1; -; Genomic DNA.
 SQ SEQUENCE 1230 AA; 136662 MW; 64B24DEB3001351 CRC64;

Alignment Scores:

Pred. No.: 1.67e-65 Length: 1230
 Score: 1029.00 Matches: 218
 Percent Similarity: 70.6% Conservative: 5
 Best Local Similarity: 69.0% Mismatches: 5
 Query Match: 87.9% Indels: 88
 DB: 2 Gaps: 1

US-09-360-685C-26 (1-685) x OAPL15_HELPY (1-1230)

```

QY 1 AAAATGCGCAAAATTAAGGATTTTCAGCAGGTAAAGCAAAAGGACCTTGAAT 60
DB LysanGlyLysanLysanLysanPheSerValThrglnAlaLysSerAplLeuLysan 765
QY 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGGTAAGGTAATGATCAATCA 120
DB SerLleYasPheValIleIleAenGlnLysIleThraPlyValAenLysSer 785
QY 766 SerLleYasPheValIleIleAenGlnLysIleThraPlyValAenLysSer 785
QY 121 GCGGTATCATGTGCTAAAGCAAGGATTTTCAGTGGGTAAAGGCTTACCGCAT 180
DB AlaValSerValAlaLysAlaThrgLysAphSerGlyValGlnAlaLeuAlaAap 805
QY 181 CTCAAAAATTTCTCAAGAGCAATGCGCCCAAGACTCAAAAAATGAATCTCAAT 240
DB LeuLysanPheSerLysGlnLysanLysanLysanLysanLysanLysan 825
QY 806 LeuLysanPheSerLysGlnLysanLysanLysanLysanLysanLysan 825
QY 181 CTCAAAAATTTCTCAAGAGCAATGCGCCCAAGACTCAAAAAATGAATCTCAAT 240
DB 786 AlaValSerValAlaLysAlaThrgLysAphSerGlyValGlnAlaLeuAlaAap 805
QY 241 GCTGAAAAAATCTGAAATATATCAATCGTTAAAGTGGTGAATGAACCTTATGC 300
DB 826 ValGlyLysLysSerGlnLysIleThraSerValLysanGlyValAenGlyThraLeuAl 845
QY 826 ValGlyLysLysSerGlnLysIleThraSerValLysanGlyValAenGlyThraLeuAl 845
QY 301 GGTATGCGGTATTCGAAAGCAAGCAACTTTTCTAAAACTTTTGGACATCAG 360
DB 846 GlyAenGlyLysSerLysValGlnAlaLysLysPheProLeuLysanLysanLysan 865
QY 846 GlyAenGlyLysSerLysValGlnAlaLysLysPheProLeuLysanLysanLysan 865
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATCAATTAAGTAAATGATCAATCAAAAAC 420
DB 866 LysGlnLysanLysanLysanLysanLysanLysanLysanLysanLysan 885
QY 421 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGGCAAGCAAGCTTGAAGA 480
DB 886 GlnProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 905
QY 481 CCGATTACCTCAAGTTGCTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 540
DB 906 ProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 925
QY 541 GCAAGTGGTTGGTGTGTAGGCGCAAGCCGGC----- 576
DB 926 AlaSerGlyLysGlnLysValGlyGlnAlaLysLysPheProLeuLysanLysanLysan 945

```



```

QY 576 ----- 576
Db 946 ValAspAspLeuSerIysValGIArgSerValSerProGluProIleTyrAlaGlnVal 965
QY 576 ----- 576
Db 966 AlaIylsValAsnAlaIylsIleAspArgLeuAsnGlnIleAlaSerGlyLeuGlyGly 985
QY 576 ----- 576
Db 986 ValGIyGlnAlaIlaGlyPheProLeuIysArgHisAspIysValAspAspLeuSerIys 1005
QY 576 ----- 576
Db 1006 ValGIyArgSerValSerProGluProIleTyrAlaThrIleAspAspLeuGIyGlyPro 1025
QY 577 TTCCTTTGAAAAAGCATGATAAGTTGATGATCTCATAGGAGGCTTTTCAGGAAT 636
Db 1026 PheProLeuIysArgHisAspIysValAspAspLeuSerIysValGIyLeuSerArgAsn 1045
QY 637 CAGGAATTGGCTCAGAAAAATTGCATCTCATCAACGCGTATCAGAA 684
Db 1046 GlnIysLeuAlaGlnIylsIleAspAsnLeuAsnGlnAlaValSerGlu 1061

```

RESULT 6

O8RNUO_HELPY

O8RNUO_HELPY PRELIMINARY; PRT; 267 AA.

```

ID O8RNUO_HELPY PRELIMINARY; PRT; 267 AA.
AC O8RNUO;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytotoxin-associated antigen (Fragment).
GN Name-caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH-CTX1;
RX MEDLINE=20123532; PubMed=10660136;
RA Garcia-de la Guardia R., Urra S., Venegas A.;
RT "Serological response to Helicobacter pylori recombinant antigens in
RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
RT gastric cancer.";
RL APMIS 107:1069-1078 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH-CTX1;
RX MEDLINE=22121498; PubMed=12125208;
RA Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
RA Bruce E., Mancilla M., Valenzuela P., Yudelevich A.;
RT "Cloning and comparison of ten gene sequences of a Chilean H. pylori
RT strain with other H. pylori strains revealed higher variability for
RT vacA and caga virulence factors.";
RL Biol. Res. 35:67-84 (2002).
DR EMBL; AF479032; AAL86902.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1
FT NON_TER 267
SQ SEQUENCE 267 AA; 29032 MW; 04B4F51480960E8B CRC64;

```

Alignment Scores:

```

Pred. No.: 3,3e-65 Length: 267
Score: 1025.50 Matches: 216
Percent Similarity: 83.4% Conservative: 5
Best Local Similarity: 81.5% Mismatches: 7
Query Match: 87.6% Indels: 37
DB: 2 Gaps: 3

```

US-09-360-685C-26 (1-685) x O8RNUO_HELPY (1-267)

```

QY 1 AAAAAAGCAAAATAGGATTTCAAGCAAGGTACCAAGCAAAAGCACTTGAAAT 60
Db 3 LysAsnGIyLysAsnIysAspPheSerIysValThrGlnAlaIylsSerAspLeuGluAsn 22
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAGCAAGGATTAATGTAATCAATCA 120
Db 23 SerIleLysAspValIlePheAsnGlnIylsIleThrAspIysValAspAspLeuAsnGln 42
QY 121 GCGGTATCAGTGGCTTAAGCAACGGGTGATTTCAAGAGGTAGAGCAAGCTTACCGCAT 180
Db 43 AlaValSerValAlaIylsAlaThrIleAspPheSerArgValGlnGlnIleAlaValAsp 62
QY 181 CTCAAAAATTTCTCAAGGAGCAATTGCGCCCAACAGCTCAAAAAATGAAGTTCAT 240
Db 63 LeuIysAsnPheSerIylsGlnIleuAlaGlnIlaGlnIylsAsnGIySerLeuAsn 82
QY 241 GGTAGAAAAAATCTGAAATATATCAATCCCTTAAGATGCTGAATGGAACCTAGTC 300
Db 83 AlaGIyLysIylsSerGlnIleTyrAlaIylsValSerIylsValSerGIyThrIleuVal 102
QY 301 GGTAAATGGTATCTCAAGCAGACCAACCACTCTTCTTAAAACTTTTCGACATCAAG 360
Db 103 GIyAsnGIyLeuSerGlnIlaGlnIlaThrThrIleuSerIylsAsnPheSerAspIleIys 122
QY 361 AAAGAGTGAATGCAAACTT---GGAATTTCAATTAACATTAATTAAGTCACTCAA 417
Db 123 LysGIuLeuAsnAlaIylsLeuPheGlnPheAsnAsnAsnAsnAsnGIyLeuIys 142
QY 418 AAC-----GAACCACTTATCTTAAGTATTAATAAAAGCAAGGCAAGCAGTAC 471
Db 143 AsnSerThrGlnProIleTyrAlaIylsValAsnIylsIylsThrGIyGlnIlaIylsSer 162
QY 472 CTGTAAGAACCACTTACCGCTCAAGTGGCTAAAGGTAAGCAAAATTAACGACATC 531
Db 163 ProGluGluProIleTyrThrGlnIlaIylsIylsValAsnAlaIylsIleAspArgLeu 182
QY 532 AATCAAAATGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 576
Db 183 AsnGlnIleAlaSerGIyLeuGIyValGIyValGIyValAlaGIyPheProLeuIysArg 202
QY 576 ----- 576
Db 203 HisAspIysValAspAspLeuSerIysValGIyArgSerValSerProGluProIleTyr 222
QY 577 ----- 577
Db 223 AlaThrIleAspAspLeuGIyGlyProPheProLeuIysArgHisAspIysValAspAsp 242
QY 610 CTCAGTAAGTAGGCTTTCAAGATCAAGATTCAGAAATGGCTCAGAAATTCAGATCTCAAT 669
Db 243 LeuSerIysValGIyLeuSerArgAsnGlnIleuAlaGlnIylsIleAspAsnLeuAsn 262
QY 670 CAGCGGTATCAGAA 684
Db 263 GlnAlaValSerGlu 267

```

RESULT 7

O52GZ6_HELPY

O52GZ6_HELPY PRELIMINARY; PRT; 1142 AA.

```

ID O52GZ6_HELPY PRELIMINARY; PRT; 1142 AA.
AC O52GZ6;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Caga (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2806;

```

RK PubMed=1597230; DOI=10.1073/pnas.0409873102;
 RA Brandt S., Kwok T., Hartig R., Konig W., Backert S.;
 RT "NF-(kappa)B activation and potentiation of proinflammatory responses
 RT by the Helicobacter pylori CagA protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
 DR EMBL; DQ011619; AY18597.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; CagA.
 DR InterPro; IPR004355; IVSec_CagA.
 DR Pfam; PF03507; CagA; 1.
 DR PRINTS; PRO1553; TYPE4SSCAG.
 FT NON_TER 1142 1142
 SQ SEQUENCE 1142 AA; 128233 MW; E53DBE9756BA29BE CRC64;

Alignment Scores:

Pred. No.:	1,43e-64	Length:	1142
Score:	1016.00	Matches:	203
Percent Similarity:	92.1%	Conservative:	7
Best Local Similarity:	89.0%	Mismatches:	18
Query Match:	86.8%	Indels:	0
DB:	2	Gaps:	0

US-09-360-685C-26 (1-685) x Q52GZ6_HELPY (1-1142)

```

QY 1 AAAAAAGCAAAATGAGATTTTCAGCAAGTAAAGCAAAAGCAAAAGCACTTGAAAT 60
DB |||||||
QY 745 LysenolyllysahenlyshapheSerlyvalThrlnalalyserValpLeuGluasn 764
DB |||||||
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGTAAAGGATTAATTCATCA 120
DB |||:::|
QY 765 SerlelyshapVallellelshenglnlyleThrshpValshpshenleuasn 784
DB ||:::|
QY 121 GCGGTATCGTGGCTAAAGCAAGCGGTAAATTCAGTAAAGCAAGCGGTAGCCGAT 180
DB ||:::|
QY 785 AlavalSerlleAlalyGlyThrlyshpsheserlyValGlnGlnAlaleuAlshp 804
DB ||:::|
QY 181 CTCMAAAATTTCTCAAGAGCAAGCAATGGCCCAACAGCTCAAAAATGAATGTCAT 240
DB |||
QY 805 LeulysahenpsheserlyGlyGlnleuAlaglnlnlAglnlyshenglnshapheasn 824
DB |||
QY 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAAGATGTCGTAAGTGAACCTTAGTC 300
DB |||
QY 825 ThrGlylyshahenSerGlnleuThrlnalalyserValshpGlyValshahenThrleuVal 844
DB |||
QY 301 GGTATGCGGTATCTCAAGCAAGCAAGCAAGCTTTTCAAAAGCTTTGCAAGTCAAG 360
DB |||
QY 845 GlyshenglyleuSerGlyleGlnAlathralaleuThrlyshahenpsheserlylelysh 864
DB |||
QY 361 AAGAGTTGAATGCAAAAGTGAATTTCAATTAACATTAACATTAATGAATGCAAAAG 420
DB |||
QY 865 LysGlnleuasnGlylyshleuLysahenpshahenshahenahenglyleuLysahen 884
DB |||
QY 421 GAACCAATTTATGCTAAAGTAAAGTAAAGCAAGGCGGCAAGCAAGCTGACCTGAAGA 480
DB |||
QY 885 GluproilelyrThrlnalalyshValshenlyshThrGlylnalalyserSerleuGln 904
DB |||
QY 481 CCCATTAGCGTCAAGTGTCTAAAGGTAAGCAAAATGAAGCAAGCTCAATCAATA 540
DB |||
QY 905 ProilelyrThrlnalalyshValshenlyshValshahenlyshleuShapshleuasnlnle 924
DB |||
QY 541 GCAAGTGTGGTGGTGTGTAAGGCAAGCAAGCGGCTTCCTTTGAAAAGGCAATGA 600
DB |||
QY 925 AlaserGlyleuGlnlyshenValGlylnlnalalyshpshleuLysahenlysh 944
DB |||
QY 601 GTTGAATGTCAGTAAAGTAAAGCTTTCAAGCAATCAAGAAATTTGCTCAAAAATTTGAC 660
DB |||
QY 945 ValshapshleuSerlyshValGlyleuSerlyshahenglnleuAlaglnlyshleap 964
DB |||
QY 661 AATTCATCAAGCGGTATCAAG 684
DB |||
QY 965 AenleuasnGlnAlavalserGln 972
DB |||

```

RESULT 8

Q8KZH3_HELPY
 ID Q8KZH3_HELPY PRELIMINARY; PRT; 1180 AA.
 AC Q8KZH3;
 RC 01-OCT-2002 (TREMblrel. 22, Created)
 DR 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DR 01-FEB-2005 (TREMblrel. 29, Last annotation update)
 DE CagA (Cag pathogenicity island protein).
 GN Name=cagA; Synonyms=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=210;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=OK112;
 RX MEDLINE=22295025; PubMed=12391297; DOI=10.1073/pnas.222375399.
 RA Hlgoshi H., Teutsuhl R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
 RA Hatakeyama M.;
 RT "Biological activity of the Helicobacter pylori virulence factor CagA
 RT is determined by variation in the tyrosine phosphorylation sites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433(2002).
 RN (2)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=OK112;
 RA Yamakawa A., Yamazaki S., Azuma T.;
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=OK112;
 RA Azuma T., Yamakawa A., Yamazaki S.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB090089; BAC10434.1; -; Genomic DNA.
 DR EMBL; AB120425; BADI4045.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; CagA.
 DR InterPro; IPR004355; IVSec_CagA.
 DR Pfam; PF03507; CagA; 1.
 DR PRINTS; PRO1553; TYPE4SSCAG.
 SQ SEQUENCE 1180 AA; 130973 MW; 623474AF2D635020 CRC64;

Alignment Scores:

Pred. No.:	4.53e-64	Length:	1180
Score:	1009.00	Matches:	207
Percent Similarity:	82.4%	Conservative:	9
Best Local Similarity:	79.0%	Mismatches:	12
Query Match:	86.2%	Indels:	34
DB:	2	Gaps:	1

US-09-360-685C-26 (1-685) x Q8KZH3_HELPY (1-1180)

```

QY 1 AAAAAAGCAAAATGAGATTTTCAGCAAGTAAAGCAAAAGCAAAAGCACTTGAAAT 60
DB |||||||
QY 750 LysahenglyshahenlyshapheSerlyvalThrlnalalyserThrshpshengln 769
DB |||||||
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGTAAAGGATTAATTCATCA 120
DB |||
QY 770 SerllelyshapAlallellelshenglnlyleThrshpshValshpshleuasnln 789
DB |||
QY 121 GCGGTATGAGTGTCTAAAGCAAGCGGTAAATTTCAATGAGTGAAGCAAGCGGTAGCCGAT 180
DB |||
QY 790 AlavalserValAlalyserlleAlaglyshapsheserlyleGlnGlnAlaleuAlshp 809
DB |||
QY 181 CTCMAAAATTTCTCAAGAGCAAGCAATGGCCCAACAGCTCAAAAATGAAGTCTCAAT 240
DB |||
QY 810 LeulysahenpsheserlyshGlnleuAlaglnlnlAglnlyshahenlyshpshasn 829
DB |||
QY 241 GGTAAAGGATTAATGCAAAATGATCAATCGTTAAAGATGTCGTAAGTGAAGCAAGCTTAGTC 300
DB |||
QY 830 ThrGlylyshahenSerlyshleuThrlnalalyserValshahenglyThrleuVal 849
DB |||
QY 301 GGTATGCGGTATCTCAAGCAAGCAAGCAAGCAAGCTTTTCAAAAATTTTGGACATCAAG 360
DB |||
QY 850 GlyshenglyleuSerlyshleuGlnAlathralaleuThrlnalalyserlyshahenpsheserlysh 869
DB |||

```


DR InterPro: IPR004355; IYSec_caga.
 DR Pfam: PF03507; Caga; 2.
 DR PRINTS: PRO1553; TYPB4SCAGA.
 SQ SEQUENCE 1179 AA; 131297 MW; 79A112P58B749787 CRC64;

Alignment Scores:

Pred. No.: 1,12e-63 Length: 1179
 Score: 1003.50 Matches: 209
 Percent Similarity: 81.7% Conservative: 5
 Best Local Similarity: 79.8% Mismatches: 13
 Query Match: 85.7% Indels: 35
 DB: 2 Gaps: 2

US-09-360-685C-26 (1-685) x Q9L5X9_HELPY (1-1179)

```

QY 1 AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTAACGCAAGCAAAAAGCACTTGAAT 60
Db 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGlnAsn 769
QY 61 TCGTTAAAGATGATCATCATCAAAAGTAAGCATTAAGTGAATTCATCA 120
Db 770 SerValLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuAsnGln 789
QY 121 GCGGTATCATGTCGCTAAAGCAACGCGGTGATTCAGTNGG3TNGAGCAACGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysAlaThrGlyAspPheSerArgValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTTCTCAAGAGCAATTTGCGCCACACAGCTCAAAAATGAAGTCAAT 240
Db 810 LeuLysAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATG3TGTGATGAACCTTAGTC 300
Db 830 ThrGlyLysAsnSerAlaLeuTyrgInSerValLysAsnGlyValAsnGlyThrLeuVal 849
QY 301 GGTATG3GTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGCGACATCAAG 360
Db 850 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspLysLeu 869
QY 361 AAAGATTAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAATGATCAAAAAC 420
Db 870 LysGlnLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnAsnGlyLeuLysAsn 889
QY 421 GAACCATTTAAGTAAAGTTAATAAAAGAAAGCAAGCGCAACGCTAGTGAAGA 480
Db 890 GlnProIleTyrgAlaGlnValaLysLysLysLysLysLysLysLysLysLysLysLysLys 909
QY 481 CCGATTACGTCAGTTCGTTAAAGGTAATTCGAAAATTCGAAATTCGAAATTCGAAAT 540
Db 910 ProIleTyrgAlaGlnValaLysLysLysLysLysLysLysLysLysLysLysLysLys 929
QY 541 GCAAGTGTGGTGGTGTGTAAGGCAACGCGGCTTCCTTTGAAAAGCATGATTA 600
Db 930 AlaSerGlyLeuGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 948
QY 601 GTTGAAT----- 606
Db 949 ValAspAspLeuSerLysValGlyArgSerValSerProGlnProIleTyrgAlaThrIle 968
QY 607 -----GATTCAGTAG 618
Db 969 AspAspLeuGlyGlyProPheProLeuLysLysIleAlaLysValaGlnAspLeuSerLys 988
QY 619 GTAGGCTTTCAAGAAATTCGTAAGTTCGAAAATTCGAAATTCGAAATTCGAAATTCGAA 678
Db 989 ValGlyLeuSerArgGlnGlnGlnLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 1008
QY 679 TCAGAA 684
Db 1009 SerGln 1010

```

RESULT 11
 Q5ZGZ5_HELPY

ID Q5ZGZ5_HELPY PRELIMINARY; PRT; 1186 AA.
 AC Q5ZGZ5;
 DT 13-SEP-2005 (Tremblrel. 31, Created)
 DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
 DE Caga (Fragment).
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NX NCBI_TaxID=210;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=P310;
 RX PubMed=15972330; DOI=10.1073/pnas.0409873102;
 RA Brandt S.; Kwok T.; Hartig R.; Konig W.; Backert S.;
 RT "NF-(kappa)B activation and potentiation of proinflammatory responses
 by the Helicobacter pylori Caga protein".
 RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
 DR EMBL: DQ011620; RefSeq: NC_019534; GenBank: F019534; GenBank: F019534;
 DR GO: GO:0019534; F019534; F019534; F019534;
 DR InterPro: IPR004355; IYSec_caga.
 DR Pfam: PF03507; Caga; 1.
 DR PRINTS: PRO1553; TYPB4SCAGA.
 FT NON_TER 1186
 SQ SEQUENCE 1186 AA; 131865 MW; CD5AC9D1DA41BF3F CRC64;

Alignment Scores:

Pred. No.: 7.47e-63 Length: 1186
 Score: 992.00 Matches: 212
 Percent Similarity: 80.7% Conservative: 5
 Best Local Similarity: 78.8% Mismatches: 10
 Query Match: 84.7% Indels: 42
 DB: 2 Gaps: 3

US-09-360-685C-26 (1-685) x Q5ZGZ5_HELPY (1-1186)

```

QY 1 AAAAAATGCAAAAATTAAGATTTTCAGCAAGGTAACGCAAGCAAAAAGCACTTGAAT 60
Db 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGlnAsn 769
QY 61 TCGTTAAAGATGATCATCATCAAAAGTAAGCATTAAGTGAATTCATCA 120
Db 770 SerValLysAspValIleIleAsnGlnLysIleThrAspLysValaLysAsnLeuAsnGln 789
QY 121 GCGGTATCATGTCGCTAAAGCAACGCGGTGATTCAGTNGG3TNGAGCAACGTTAGCCGAT 180
Db 790 AlaValSerValAlaLysAlaThrGlyAspPheSerArgValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTTCTCAAGAGCAATTTGCGCCACACAGCTCAAAAATGAAGTCAAT 240
Db 810 LeuLysAsnPheSerArgGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 829
QY 241 GGTATG3GTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGCGACATCAAG 300
Db 830 ThrGlyLysAsnSerAlaLeuTyrgInSerValLysAsnGlyValaAsnGlyThrLeuVal 849
QY 301 GGTATG3GTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGCGACATCAAG 360
Db 850 GlyAsnGlyLeuSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 869
QY 361 AAAGATTAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAATGATCAAAAAC 420
Db 870 LysGlnLeuAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 888
QY 421 -----GATTCAGTAG 429
Db 889 GlyLysAspLysGlyProGlnGlnProIleTyrgAlaLysValaLysLysLysLysLysLys 908
QY 460 CAAGCAGTTCGTTGAAGAACCAATTCGTAAGTTCGAAAATTCGAAATTCGAAATTCGAA 519
Db 909 GlnAlaLysSerProGlnGlnProIleTyrgAlaGlnValaLysLysLysLysLysLysLys 928

```

```

QY      520 ATTGACCGACTCAATCAATAATAGCAATGGTTGGGTGGTGAAGGGCAGCACGCCGC--- 576
Db      925 IIEBPAHGLEUENMYSIIIEALASERTGLYLEUGLYGLYVALGLYNIAIAGLIPHE 948
QY      576 ----- 576
Db      949 ProLeuLyHisAspLysValAspAspLeuSerLysValGIlyArGerValSerPro 968
QY      577 -----TTCCCTTGTAAGAAGCATGAT 597
Db      969 GIUProileTyRAlaThrIleAspAspLeuGIlyGIlyProPheProLeuLysLysHisAsp 988
QY      598 AAAGTGTGATGATCTCAGTAAGGTAGGGCTTCAGAAACAACGAATTGGCTCAGAAAT 657
Db      988 LysValAspAspLeuSerLysValGIlyLeuSerArgAsnGlnLysLeuAlaGlnLysIle 1000
QY      658 GACAATTCATCAATCAACGGGTATCAGAA 684
Db      1009 AspAsnLeuAsnGlnAlaValSerGlu 1017

RESULT 12
Q9F222_HELPY                                PRT; 1179 AA.
ID   Q9F222_HELPY                             PRELIMINARY;
AC   Q9F222;
DT    01-MAR-2001 (TrEMBLrel. 16, Created)
DI    01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
PT    01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE    Caga.
GN     Name=caga;
OS    Helicobacter pylori (Campylobacter pylori).
OC    Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
CX    Helicobacteraceae; Helicobacter.
NCBI_TaxId=210;
RN     [1]
RP     NUCLEOTIDE SEQUENCE.
RC     STRAIN=ATCC43579;
RX     MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s005350070002;
RA     Hosho H, F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA     Ando T.;
RT     "Heterogeneity found in the caga gene of Helicobacter pylori from
RT     Japanese and non-Japanese isolates.";
RL     J. Gastroenterol. 35:890-897(2000).
DR     EMBL; AB015414; BAB20924.1; -, Genomic DNA.
DR     GO; GO:0019534; F:toxin transporter activity; IEA.
DR     InterPro; IPRO05169; Caga.
DR     pfam; PF03507; Caga_1.
DR     PRINTS; PRO1553; TYPE4SSCAG.
SQ     SEQUENCE 1179 AA; 131686 MW; 7D0ADC34B6446434 CRC64;

Alignment Scores:
Pred. No.:          9,576-63                      Length:         1179
Score:              990.50                        Matches:           206
Percent Similarity: 80.8%                         Conservative:       5
Local Similarity:  78.9%                          Mismatches:        17
Query Match:       84.6%                           Indels:            33
DB:                2                               Gaps:              1

US-09-360-685C-26 (1-685) x Q9F222_HELPY (1-1179)

QY      1 AAAATGGCAAAAATATAGGATTTCAGCAAGGTACGCAAGCAAAAAAGCAGCTTGAAT 60
Db      750 LysAsnGlyLysAsnLysAspHisSerLysValThrGlnAlaLysSerAspLeuLysAsn 769
QY      61 TCCGTTAAGATGTATCATCATCAATCAATAAAGGTAAAGGTAAAGTATCATCTCAATCAA 120
Db      770 SerValLysAspAlaIlePheAsnGlnLysValIleAspLysValAspAspLeuAsnGln 789
QY      121 GGCGTATCAGTGGCTAAAGCAACGGGTGATTTCACTAGGGTAGAACCAAGCTTAGCCGAT 180
Db      790 AlaValSerValAlaLysAlaThrSerAspPheSerArgValGluGlnAlaLeuAlaAsp 809
QY      181 CTCAAAAATTTCTCAAGAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240

```

Alignment Scores:

```
Db      810  LeuIysaenPheSerLysGlnGlnLeuAlaGlnGlnThrGlnIlysaenGlnSerPheasn 829
QY      241  GCTAATAAAAAAATCGAATATATCAATCCCGTTAAGATGGGGAATGAAGAACCTATCTC 300
Db      830  ValGlyIyLysSerGluIleTyriGlnSerValIySaenGlyValaenGlyThrLeuVal 849
QY      301  GGTAAATGGGGTATCTCAAGCAGAAAGCCCAACTCTTTCTTAAAACTTTCCGACATCAG 360
Db      850  GlyThrGlyLeuSerLysValaGlnIleThrThrLeuSerLysaenPheSerAspIleLys 869
QY      361  AAAAGGTGATGCAGAAACTTGGAAATTTCAATATCAATATCAATATGACTCAAAAC 420
Db      870  LysGlnLeuSaenAlaLysLeuGlnLysaenPheSaenSaenSaenSaenSaenGlyLeuLysaen 889
QY      421  GAACCATTTATNGCTAAGTTAATATTAAGAAAGCAGGGCAGACGCTAGCCTGGAAGA 480
Db      890  GluProlIleTyriAlaGlnValaIasnLysLysValaGlyGlnAlaAlaSerHisGlnGlu 909
QY      481  CCCATTATACGCTCAAGTGTCTTAAAGGTAAATGCAAAATATGACCGACCTCAATCAATA 540
Db      910  ProlIleTyriThrGlnValaIalalyeLysValaIasnAlaLysIleAspGlnLeuAsnGlnAla 929
QY      541  GCAAGTGGTTTGGCGTGTGTAGGCAACA----- 570
Db      930  AlaSerGlyLeuGlyGlyAlaIaasnGlnAlaGlyPheProLeuLysArgHisAspLysVala 949
QY      570  ----- 570
Db      950  AspaPheLysSerLysValaGlyLeuSerAlaIasnHisGluProlIleTyriAlaThrIleAsp 969
QY      571  -----GCGGGGCTTCCTCTTTGAAAAGCGATGATTAAGTTGATGATCTTCAGTAAGGTA 621
Db      970  AspaLeuGlyGlyProPheProLeuLysArgHisAspLysValaAspaPheLysSerLysVala 989
QY      622  GGGCTTTAAGACATCAAGATTGGCTTCAGAAATTTGACAACTTCATCAATCAAGCGGTATCA 681
Db      990  GlyLeuSerArgaSpGlnGlnLeuAlaGlnLysIleAspaLeuLeuAsnGlnAlaValaSer 1009
QY      682  GAA 684
Db      1010  Gln 1010

RESULT 13
Q6VRG7 HELPY
ID      06VRG7 HELPY PRELIMINARY;          PRT; 1182 AA.
AC      06VRG7_
DT      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
GN      Name=caga; ORFNames=HP0547;
OS      Helicobacter pylori (Campylobacter pylori).
OC      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC      Helicobacteriaceae; Helicobacter.
OX      NCBI_TaxID=210;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=DU23.2;
RA      PubMed=15019587; DOI=10.1016/j.gene.2003.11.029;
RA      Blomstergren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
RT      "Comparative analysis of the complete caga pathogenicity island
RL      sequence in four Helicobacter pylori isolates.";
RL      Gene 328:85-93(2004).
DR      EMBL, AY330644; AAR03970.1; -; Genomic DNA.
DR      GO; GO:0019534; F:toxin transporter activity; IEA.
DR      InterPro; IPR005169; Caga.
DR      InterPro; IPR004355; IVSec_caga.
DR      Pfam; PF03507; Caga_1.
DR      PRINTS; PRO1553; TYPEIASSCAGA.
SQ      SEQUENCE 1182 AA; 131615 MW;  BDCC536BBAF1E27 CRC64;
```

Pred. No.: 2,796-62 Length: 1182
Score: 994.00 Matches: 207
Percent Similarity: 80.3% Conserved: 5
Best Local Similarity: 78.4% Mismatches: 16
Query Match: 84.0% Indels: 36
Gaps: 2

US-09-360-685C-26 (1-685) x Q6VRG7_HELPY (1-1182)

```
QY 1 AAAAATGCGAAATAATAGATTTCACGAGTAAACGACGCAAAAAGGACCTTGAAAT 60
    |||||
Db 750 LysAmGlyLysAmLysAspPheSerLysValThrGlnLalysSerAspLeuGln 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAAGTAAAGGATTAATGATTCATCAATCA 120
    |||||
Db 770 SerValLysAspValLleLleAmGlnLysValThrAspLysValAspAmLeuAmGln 789
QY 121 GCGGTATCAGTGGCTAAAGCAAGGCTGATTTCAGTGGGTAGAGCAAGGCTTAGCGCAT 180
    |||||
Db 790 AlavaSerLleAlaLysAlaLleGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCTCAAAAAATGAAAGTCCAT 240
    |||||
Db 810 LeuLysAmPheSerLysGlnGlnLeuAlaGlnGlnLalysAmGlnAspPheAm 829
QY 241 GCTAGAAAAATCTGAAATATATCAATCCGTTAAGATGCTGTAATGAAACCTTAGTC 300
    |||||
Db 830 ThrGlyLysLysSerGlnLleLysGlnSerValLysAmGlyValAmGlyThrLeuVal 849
QY 301 GGTATGGGTTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGGACATCAG 360
    |||||
Db 850 GlyAmGlyLysSerGlyLleGlnLalThrAlaLeuAlaLysAmPheSerAspLleLys 869
QY 361 AAAGATTTGATGCAAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAAC 420
    |||||
Db 870 LysGlnLeuAmGlnLysPheLysAmPheAmGlnAmGlnAmGlnLysLeuGlnAm 889
QY 421 -----GAACCCATTTATGCTAAAGTAAATTAATTAAGAAAGCAAGGCAAGCTAGCTT 474
    |||||
Db 890 SerThrGlnProLleLysAlaLysValAlaLysLysLysThrGlnAlaLalSerPro 909
QY 475 GAAGAAACCATTTACGCTCAAGTTGCTTAAAGGTAATGCAAAAATTTGACGACAT 534
    |||||
Db 910 GlnGlnProLleLysAlaGlnValAlaLysLysValAlaLysLleAspLleAsp 929
QY 535 CAATTAAGCAAGTGGTGGTGGTGGTGGGCAAGCAAGCGGCTTCCCTTGAAGAGGAT 594
    |||||
Db 930 GlnLleAlaSerGlyLeuGlyGlyValGlyGlnAlaValGlyPheProLeuLysAlaGln 949
QY 595 GAT----- 597
    |||||
Db 950 AspLysValAspAspLeuSerLysValGlyArgSerValSerProGlnProLleLysAla 969
QY 598 -----AAAGTTAGATGATC 612
    |||||
Db 970 ThrLleAspLeuGlyGlyProPheProLeuThrGlnLysSerLysValAspLeu 989
QY 613 AGTAAGTAAGGCTTCAAGCAATCAAGAAATGGCTCAAAAATTTGACATTCATCA 672
    |||||
Db 990 SerLysValGlyLeuSerLysAspGlnLysLeuAlaGlnLysLleAspAmLeuAmGln 1009
QY 673 GCGGTATCGAA 684
    |||||
Db 1010 AlavaSerGln 1013
```

RESULT 14

Q6VRG7_HELPY
ID Q6VRG7_HELPY PRELIMINARY; PRT; 394 AA.

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cytotoxin associated protein Caga (Fragment).

GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
NC NCBI_TaxID=210;
NM [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC51407;
RA Yamoka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,
RU Ramirez F.C., Mahachai V., Oato M.S., Graham D.V.,
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB057003; BAB87370.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1 394
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43035 MW; BIC9D5596ABE6EBF CRC64;

Alignment Scores:

Pred. No.: 3,86e-62 Length: 394
Score: 982.50 Matches: 211
Percent Similarity: 76.9% Conserved: 5
Best Local Similarity: 75.1% Mismatches: 10
Query Match: 83.9% Indels: 55
Gaps: 3

US-09-360-685C-26 (1-685) x Q6VRG7_HELPY (1-394)

```
QY 1 AAAAATGCGAAATAATAGATTTCACGAGTAAACGACGCAAAAAGGACCTTGAAAT 60
    |||||
Db 69 LysAmGlyLysAmLysAspPheSerLysValThrGlnLalysSerAspLeuGln 88
QY 61 TCCGTTAAAGATGATCATCAATCAAAAAGTAAAGGATTAATGATTCATCAATCA 120
    |||||
Db 89 SerValLysAspValLleLleAmGlnLysLleThrAspLysValAspAmLeuAmGln 108
QY 121 GCGGTATCAGTGGCTAAAGCAAGGCTGATTTCAGTGGGTAGAGCAAGGCTTAGCGCAT 180
    |||||
Db 109 GlnValLeuValAlaLysAlaLysThrGlyAspPheSerMetValGlnGlnAlaLeuAlaAsp 128
QY 181 CTCAAAAATTTCTCAAGCAAGCAATTTGCCCCAAGCTTCAAAAAATTTGAACTCAT 240
    |||||
Db 129 LeuAmAmPheSerLysGlnGlnLeuAlaGlnGlnLalysAmGlnLysLeuAm 148
QY 241 GCTAGAAAAATCTGAAATATATCAATCCGTTAAGATGCTGTAATGAAACCTTAGTC 300
    |||||
Db 149 AlaglyLysLysSerGlnLleLysGlnSerValLysAmGlyValAmGlyAlaLeuVal 168
QY 301 GGTATGGGTTATCTCAAGCAAGCAACCACTTTCTTAAAACTTTTGGACATCAG 360
    |||||
Db 169 GlyAmGlyLysSerGlnAlaGlnLalThrLleSerAspAmLeuSerAspLleLys 188
QY 361 AAAGATTTGATGCAAACTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAAC 420
    |||||
Db 189 LysGlnLeuAmLalysLysLeuLysAmPheAmAmAmAmArg-----LeuLysAm 206
QY 421 -----GAA 423
    |||||
Db 207 GlnProLleLysAlaLysValAmLysLysValAlaGlyGlnAlaLalSerProGlnLle 226
QY 424 CCCATTTATGCTAAAGTAAATTAATTAAGAAAGCAAGGCAAGGCTAGCTTGAAGAACCC 483
    |||||
Db 227 ProLleLysAlaLysValAmLysLysValAlaGlyGlnAlaLalSerProGlnLlePro 246
QY 484 ATTACGCTCAAGTGGCTTAAAGGTAATGCAAAAATTTGACGACATTCATCAATGCA 543
    |||||
Db 247 LleLysAlaGlnValAlaLysLysValAlaLysLleAspLleAspAmLeuAmGlnLle 266
QY 544 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 603
    |||||
Db 267 SerGlyLeuGlyValAlaGlyGlnLalSerGlyPheProLeuLysAlaGlnLysVal 286
QY 604 GATGATCTCACT----- 615
```

```

Db      287 AspAsnLeuSerLysValGlyLeuSerAlaArgSerGluProIleTyrAlaThrIleAsp 306
Qy      616 -----AAAGTA 621
Db      307 GlyLeuGlyGlyProIleProLeuLysArgHisAspLysValAspGluLeuIleLysVal 326
Qy      622 GGGCTTTCAGAGATCAAGATTCGTCAGAAATTCAGATTCATCAATCAAGCGGTATCA 681
Db      327 GlyLeuSerArgThrGlnLysLeuAlaGlnLysIleAspAsnLeuAsnGlnAlaValSer 346
Qy      682 GAA 684
Db      347 Glu 347

RESULT 15
08KZB0_HELPY
ID 08KZB0_HELPY PRELIMINARY; PRT; 327 AA.
AC 08KZB0;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Caga (Fragment).
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RC MEDLINE=22295025; Pubmed=12391297; DOI=10.1073/pnas.222375399;
RA Higashi H., Tsutsumi R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
RA Hatakeyama M.,
RT "Biological activity of the Helicobacter pylori virulence factor Caga
RT is determined by variation in the tyrosine phosphorylation sites.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433(2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RA Yamakawa A., Yamazaki S., Azuma T.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB090151; BAC10497.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1
FT NON_TER 327
SQ SEQUENCE 327 AA; 35798 MW; 1309C5E2418C7DEC CRC64;

```

Alignment Scores:

Pred. No.:	2.03e-61	Length:	327
Score:	972.50	Matches:	197
Percent Similarity:	90.8%	Conservative:	10
Best Local Similarity:	86.4%	Mismatches:	20
Query Match:	83.0%	Indels:	1
DB:	2	Gaps:	1

US-09-360-685C-26 (1-685) x 08KZB0_HELPY (1-327)

```

Qy      1 AAAAATGCAAAAATAGGATTTCAGCAAGGTAAAGCAAAAGCGACCTTGAAT 60
Db      59 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGlnVal 78
Qy      61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAACGATTAAGTTGATTAATCTCAATCAA 120
Db      79 SerIleLysAspValIleIleAsnGlnGluIleThrAspLysValAspAsnLeuAsnGln 98
Qy      121 GGGGTATCAAGTCTAAAGCAAGCGGTGATTTCAGTAGGTAAAGCAAGGTTAGCCGAT 180
Db      99 AlaValSerIleAlaAsnSerThrGlyAspPheSerGlyValGlnGlnAlaLeuAlaAsp 118
Qy      181 CTCAAAATTTCTCAAGAGCAATGGCCCAACAAGCTCAAAAATATGAATGATCTCAAT 240

```

```

Db      119 LeuLysAsnPheSerLysGlyGlnLeuThrGlnAlaGlnLysAsnGlnAspPheAsn 138
Qy      241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGCTGATGAGACCTTAGTC 300
Db      139 ThrGlyLysAsnSerGluLeuTyrGlnSerValLysAsnGlyValAsnGlyThrLeuVal 158
Qy      301 GGTAAATGGGTATCTCAAGACAGAACCAACCTTCTTAAAAACCTTTCCGACATCAAG 360
Db      159 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrLeuSerLysAsnPheSerAspIleLys 178
Qy      361 AAAGAGTTGAATGCAAACTTGAAATTTCAATPACATPACATTAATGACTCAAAAAC 420
Db      179 LysGluLeuAsnAlaLysLeuGlnLysAsnPheAsnAsnAsnAsnGlnLysValAsn 198
Qy      421 GAACCATTTATCTTAAAGTTAATTAAGAAAGCAAGCGCAACAGCTAGCCTTGAAAGA 480
Db      199 GluProIleTyrAlaLysValAsnLysGlnLysThrGlyGlnValAlaAsnLeuGlnLys 218
Qy      481 CCCATTAGCGCTCAAGTTGCTAAAGGTAAATGCAAAATTCAGCCGACTCAATCAATTA 540
Db      219 ProIleTyrThrGlnValAlaLysValLysAlaLysIleAspArgLeuAspGlnIle 238
Qy      541 GCAGTGTGTTGGGTGTGTAGGGCAAGCAGCGGCTTCCTTGAAGCAATGATTA 600
Db      239 AlaSerGlyLeuGlyGlyValGlnGln---AlaGlyPheSerLeuLysGlnIle 257
Qy      601 GTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATGGCTCAGAAATTA 660
Db      258 ValGlyAspLeuSerLysValGlyLeuSerArgGlnGlnLysGlnLysIleAsp 277
Qy      661 AATCTCAATCAAGCGGTATCAAGAA 684
Db      278 AsnLeuAsnGlnAlaValSerGlu 285

```

Search completed: March 6, 2006, 20:00:12
Job time : 223 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 6, 2006, 19:53:26 ; Search time 7.6 Seconds
(without alignments)
1734.434 Million cell updates/sec

Title: US-09-360-685C-26
Sequence: 1 aaaaatggcaaaaataagga.....caatcaagcgcgtatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 segs, 96216763 residues
Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODR=frame+.n2p.model -DEV=xlp
-Q=/abs/ABSSWEB.spool/US09360685/runat.06032006.121318.15398/app.query.fasta_1
-DB=PIR -QFMT=fastaan -SUFFIX=pr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -ALIGN=15 -MODR=LOCAL
-OUTFMT=prc -NORM=ext -HEAPSIEB=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p
-USER=US09360685@CCN_1_1.63@runat.06032006.121318.15398 -NCPU=6 -ICPU=3
-NO_MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WAPR TIMEOUT=30 -TREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR.80.*
2: PIR1.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1083	92.5	1215	2 B48281
2	959.5	81.9	1186	2 C64588
3	896.5	76.6	1167	2 B71924
4	140	12.0	6713	2 B89921
5	127.5	10.9	774	2 JC2299
6	125	10.7	2481	2 D90011
7	124	10.6	504	2 D71615
8	124	10.6	1302	1 JC6009
9	121	10.3	993	2 C90072
10	121	10.3	1125	2 E90598
11	121	10.3	2401	2 T28676
12	119.5	10.2	821	2 S67087
13	119	10.2	615	2 AB2641
14	119	10.2	622	2 D97423

15	118	10.1	1713	2 A55347	adhesive ligand ep
16	117.5	10.0	482	2 C86322	hypothetical prote
17	117	10.0	672	2 S61463	p83/100 protein -
18	117	10.0	693	2 S61464	p83/100 protein -
19	117	10.0	693	2 I40090	p93 protein - Lyme
20	116	9.9	667	2 B97012	methyl-accepting c
21	116	9.9	1175	2 D35815	myosin heavy chain
22	116	9.9	1175	2 C35815	myosin heavy chain
23	116	9.9	1201	2 A35815	myosin heavy chain
24	116	9.9	1201	2 B35815	myosin heavy chain
25	116	9.9	2385	2 A32491	myosin heavy chain
26	116	9.9	2411	2 B32491	myosin heavy chain
27	115.5	9.9	1269	2 F64730	probable myosin he
28	115	9.8	1191	2 B97116	chromosome segrega
29	114.5	9.8	481	2 T18465	hypothetical prote
30	114.5	9.8	1051	2 T18351	hypothetical prote
31	114.5	9.8	1120	2 A10561	integral membrane
32	114.5	9.8	1365	2 T30822	impl protein - Myc
33	114.5	9.8	1939	1 A46762	impl protein - Myc
34	114	9.7	1005	2 A64465	myosin alpha heavy
35	114	9.7	3672	2 T23433	hypothetical prote
36	114	9.7	3704	2 S18253	probable laminin a
37	114	9.7	3712	2 H97217	laminin alpha-1 ch
38	113.5	9.7	722	2 A92524	uncharacterized co
39	113.5	9.7	955	1 A35254	leukotoxin A - Pas
40	113.5	9.7	2139	2 T18296	myosin heavy chain
41	113	9.6	1039	2 S62509	probable vesicular
42	113	9.6	1044	2 T50213	mechanosensitive c
43	113	9.6	1120	2 P06693	mechanosensitive c
44	113	9.6	1120	2 B85544	prophage p13 prote
45	113	9.6	1640	2 D86798	

ALIGNMENTS

RESULT 1

B48281

Cytotoxin-associated gene A protein caga - Helicobacter pylori (strains G39G and CCUG
N/Alternate names: Immunodominant 128k antigen; immunodominant 135k antigen
C/Species: Helicobacter pylori

C/Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C/Accession: B48281; A48281

R/Covacci, A.; Gensini, S.; Bugnoli, M.; Petraccia, R.; Burroni, D.; Macchia, G.; Masso

Proc. Natl. Acad. Sci. U.S.A. 90, 5791-5795, 1993

A/Title: Molecular characterization of the 128-kDa immunodominant antigen of Helicobac

A/Reference number: A48281; MUID:93296225; PMID:8516329

A/Accession: B48281

A/Molecule type: DNA

A/Residues: 1-1215 <COV1>

A/Cross-References: UNIPROT:P80200; UNIPARC:UP1000017A916; GB:X70039; NID:G394912; PID

A/Experimental source: Strain G39G

A/Note: Sequence extracted from NCBI backbone (NCBIN:133971, NCBI:P.133973)

A/Accession: A48281

A/Molecule type: DNA

A/Residues: 1-957,1026-1215 <COV2>

A/Cross-References: UNIPARC:UP1000002P93F; GB:X70039; NID:G394912; PID:CAA49633.1; PI

A/Experimental source: Strain CCUG 17874

A/Note: Sequence extracted from NCBI backbone (NCBIN:133974, NCBI:P.133968)

C/Comment: These sequences are from cytotoxin producing strains. It is similar to a se

F:958-1025/Region: 33-Residue repeats

F:958-1025/Region: 33-Residue repeats

F:958-1025/Region: 33-Residue repeats

F:958-1025/Region: 33-Residue repeats

US-09-360-685C-26 (1-685) x B48281 (1-1215)

QY 1 AAAAATGGCAAAAATTAAGATTTCAGCAGAGTAACGACGCAAAAGCCGACTTGAAAT 60

Db 750 LysAenGlyLysAenLysAenPheSerLysValThrClnalLysSerAenLysGluAen 769

QY 61 TCGGTAAAGATGTGATCATCAATCAAAAGCTAAGCATTAAGTTGATATTCATCA 120

Db 770 SerValLysAenValIleIleAenGlnLysValIleAenLysValAenAenGln 789

QY 121 GCGGTATCAGTGGCTAAAGCAACGGCTGATTTCACTAGAGGTAGAGCAACGGTTAGCCCAT 180

Db 790 AlaValSerValAlaLysAlaThrGlyAenPheSerArgValGlnAlaIleValAaP 809

QY 181 CTCAAAAATTTCTCAAGAGACATTTGGCCCAACAGCTCAAAAAATGAAATGCTCAAT 240

Db 810 LeuLysAenPheSerLysLeuGlnLysAlaGlnGlnAlaGlnLysAenGlnSerLeuAen 829

QY 241 GGTGAAAAAATCTGAAATATATCAATCCGTAAAGATGGTGAATGAAACCCCTAGTC 300

Db 830 AlaAlaGlyLysSerLeuIleTyrGlnSerValLysAenGlyValAenGlyThrLeuVal 849

QY 301 GGTATGGCTTATCTCAAGCAAGCAACCTCTTCTTAAAACTTTTGGCAATCAAG 360

Db 850 GlyAenGlyLysSerGlnAlaGlnAlaThrThrLeuSerLysAenPheSerAenLys 869

QY 361 AAAGAATTGAATGCAAACTTGGAAATTTCAATTAACATTAACATTAATGACCTCAAAAC 420

Db 870 LysGlnLysAenAlaLysLeuGlyAenPheAenAenAenAenAenGlyLysAen 889

QY 421 GAACCATTTATGCTAAAGTTAATTAATAAAGAAACGGGCAAGCGCTACCTTGAAGA 480

Db 890 GluProIleTyrAlaLysValAenLysLysAlaGlyGlnAlaIleAenSerLeuGln 909

QY 481 CCCATTAGCTCAAGTTGCTAAAAAGGTAATGCAAAATTAACGCACTCAATCAATA 540

Db 910 ProIleTyrAlaGlnValAlaLysLysAlaAenAlaLysIleAenAenAenAenIle 929

QY 541 GCAAGGTGTTGGGGTGTGAGGCAAGCAACGGCGCTTCCCTTTGAAAGGCAATGAATA 600

Db 930 AlaSerGlyLeuGlyValValGlyGlnAlaIleAglPheProLeuLysArgHisAenLys 949

QY 601 GTTGAATGATCTCCATTAAGCTAAGG----- 624

Db 950 ValAenAenPheSerLysValGlyLysArgSerValSerProGluProIleTyrAlaThr 969

QY 624 ----- 624

Db 970 IleAenAenPheLysGlyLysProPheProLeuLysArgHisAenLysValAenAenLysSer 989

QY 624 ----- 624

Db 990 LysValGlyArgSerValSerProGluProIleTyrAlaThrIleAenAenLysGlyLys 1009

QY 625 -----CTTCAAGCAAT 636

Db 1010 ProPheProLeuLysArgHisAenLysValAenAenLysSerLysValLeuSerAenAen 1029

QY 637 CAAGATTGGCTCAAGAAATTAACAATTCATCAAGCGGTATTCGAA 684

Db 1030 GlnGlnLeuAlaGlnLysIleAenAenLysAenGlnAlaIleAenSerLys 1045

RESULT 2

C64588

cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)

N/AIternate names: cytoCoxin-associated gene A protein

Cispecies: Helicobacter pylori

CDate: 09-Aug-1997 #sequence_revision 09-Aug-1997 #ext_change 09-Jul-2004

CAccession: C64588

R/Tomb, J.F.; White, O.; Kesteven, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna, N.; J.D.; Kelley, J.M.; Cotton, M.D.; Wetmann, J.W.; Fujii, C.; Bowman, C.; Watthey, L. A.; Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpx, P.D.; Smith, H.O.; Fraser, C. A.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. Reference number: A65520; MUID:97394467; PMID:9252185

```

A>Accession: C64598
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A>Molecule type: DNA
A>Residues: 1-1186 <RNA>
A>Cross-references: UNIPROT:P55980; UNIPARC:UPI0000126DD8; GB:AE000569; GB:AE000511; NTD

Alignment Scores:
Pred. No.:      1,41e-61
Score:          959.50
Percent Similarity: 78.4%
Best Local Similarity: 77.7%
Query Match:    81.9%
DB:
      DB:          2
      Gaps:        3

US-09-360-685C-26 (1-685) x C64598 (1-1186)

```

OY	1	AAAAATGGCAAAAATTAAGATTTCAGCAAGCTACAGCAACGAAAAAGCACTTGAAAT	60
Db	755	LyasngILyIsanlyIsapheSerLyValThrGlnAlaIysSerAspLeuGluAen	774
OY	61	TCCGTTAAAGATGGTATCATTCATCAAAAGGTACCGATAAAGTTGATATCTCAATCA	120
Db	775	SerVallyIsapheValIleIleIenGlnLysValThrAspLyValAspAsnLeuAenGln	794
OY	121	GGCGTATCGAGGCTAAACCAACGGATATTCAAGTAGGGTAGACCAAGCTTACCGCAT	180
Db	795	AlaValSerValAlaIvalAlaMetGlyAspPheSerTrpValGluGlnValIleuAlaAsp	814
OY	181	CTCAAAATTTCTTCAAAGAGCAATTGGCCCAACAGCTCTCAAAAAATGCAATCTCAT	240
Db	815	LeuLyAsnPheSerLyGluGlnLeuAlaGlnGlnIleGlnLyAsnGluAspPheAsn	834
OY	241	GCTAAGAAAAATCTGAATATATCATTCCTGTAGAAAGGTGTGAATGCAACCTTATC	300
Db	835	ThrGlyLyAsnSerIleuLyTrpIleSerValLyAsnSerValAsnLySerIleuVal	854
OY	301	GCTAATGGGTATCTTCAAGCAAGCAACACTTCTTCAAAAATCTTTGGACATCAAG	360
Db	855	GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaIleuAlaIysAsnPheSerAspIleLys	874
OY	361	AAAGATTGATGCAAACTTGGAAATTTCAATCAATPAACATTAATGACTCAAAAAC	420
Db	875	LySLeuAsnGlnLysPheLyAsnPhe---AsnAsnAsnAsnAsnGlnLysLeuAsn	893
OY	421	-----GAACCCATTTATGCTAAAGTTAATTAATAAAGAGCGGCAACAGCTTACCTT	474
Db	894	SerThrIleuProIleTyAlaIysValIleuLysLysLysThrGlyGlnAlaIleSerPro	913
OY	475	GAGAACCCATTTAGCGCTCAAGTCTTAAAGATGAATAGCAAAATTTACCCATCTCAT	534
Db	914	GluIleuProIleTyThrGlnValAlaIysLysValAsnAlaLysIleAspArgLeuAsn	933
OY	535	CAATAGCAAGTGTGGGTGTGTAGGGCAAGCAAGCGGCTCCCTTTGAAAAGGCAT	594
Db	934	GlnIleAlaSerGlyLeuGlyGlyValGlyGlnAlaAlaGlyLysPheTrpLeuLysValGlnIle	953
OY	595	GATAAAGTTGATGATCTCAAGTAGGTAGGCTT-----	627
Db	954	AspLyValAspAspLeuSerLyValGlyLeuSerAlaSerProGluProIleTyAla	973
OY	627	-----	627
Db	974	ThrIleAspAspLeuGlyLysProPheTrpLeuLysArgIleAspLyValAspAspLeu	993
OY	628	-----TCAAGAAATCAAGAAATTTGGCTACAGAAATTTGACATCTCAATCA	672
Db	994	SerLyValAlaGlyAsnSerTrpAsnGlnLysLeuAlaGlnLysIleAspAsnLeuAenGln	1013
OY	673	GGCGTATCAAG	684
Db	1014	AlaValSerGlu	1017

cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori B71924

C.Species: Helicobacter pylori
A.Variety: strain J99
C.Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C.Accession: B71924
R.Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 387, 176-180, 1999
A.Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
A.Reference number: A71800; MUID:99120557; PMID:9923682
A.Accession: B71924
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1167 <ARN>
A.Cross-references: UNIPROT:Q9ZLT1; UNIPARC:UPI0000126DD7; GB:AE001483; GB:AE001439; NITD:NC_009616
A.Experimental source: strain J99
C.Genetics:
A.Gene: cagA

Alignment Scores:

Pred. No.:	4.72e-57	Length:	1167
Score:	896.50	Matches:	189
Percent Similarity:	74.8%	Conservative:	7
Best Local Similarity:	72.1%	Mismatches:	15
Query Match:	76.6%	Indels:	51
DB:	2	Gaps:	2

US-09-360-685C-26 (1-685) X B71924 (1-1167)

QY 1 AAAAAATGGCAAAATTAAGATTTCAGCAGGTAACGCAAAACCGACTTGAAAT 60
DB 754 LysasnnglylsbaenlysaaspheSerlysaValThrGlnalysSerAapleuGln 773
QY 61 TCCGTTAAAGATGATCATCATCAATCAAAAGGTAACGGAATAGTTGATATTCACAA 120
DB 774 SerlelysaapvallellelsanGlnlyseThrApLyValaAapleuasnGln 793
QY 121 GCGGTATCAGTGCTAAAGCAACGGGTGATTCAGTAGGGTAGAGCAAGCGTTAGCCAT 180
DB 794 AlalvalserValalalyalThrGlyAspPheSerGlyValGlnGlnalaleuAlaasp 813
QY 181 CTCAAAAATTTCTCAAGAGCAATTGGCCCAACAGCTCAAAAAATGAAAGTTCAT 240
DB 814 LeuylsaenPheSerlyGlnGlnleuAlaGlnGlnalaleuAlaaspPheasn 833
QY 241 GCTGAAGAAAAATGTGAATATATCAATCCGTTAAGATGGTGTGAATGGAACCTTAGTC 300
DB 834 ThrGlylsbaenSerAlaleuylGlnSerVallylsasnGlyValaasnGlyThrleuVal 853
QY 301 GGTATGGGTATCTCAAGCAGAACCAACACTCTTCTTAAAACTTTGGGACATCAAG 360
DB 854 GlyasnGlyleuSerlysaIaglnAlaThrThrleuSerlysaenPheSerAapllely 873
QY 361 AAAGAGTTGAATGCAAAACCTTGAAATTCATTAACAATAACAATAATGACTCAAAAC 420
DB 874 LysGlnleuAlaenAlalalyseuGlyasnPheasnleuasnleuasnleuGlnleuasn 893
QY 421 GAACCATTTATGCTAAAGTTAATAAAAAGAAACGAGGCAAGCACTAGCTTGAAGAA 480
DB 894 Ser-----ThrGln 896
QY 481 CCCATTTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGCACTCAATAA 540
DB 897 ProlelyrThrGlnValalalylyVallyalalyAlaIysleAapleuAapleuGlnlle 916
QY 541 GCAAGTGTTGGGTGTGTTAGGGCAAGCGGCG----- 576
DB 917 AlaserGlyleuGlyAapvalGlyGlnalalaserPheleuLeuylsaIGHI:saaply 936
QY 576 ----- 576
DB 937 ValaAapleuSerlysaValGlyleuSerAlaasnHleuGlnProllelyrAlaThrille 956

[illegible]

```

Db      5964 A l a l e u l y s g l u i l e l y s a p l e u v a l l y s a l a l y s g l u a s n a l a l y s g l a s p v a l a s p
Qy      361 A A A G T G A A T G C A A A C T T G A A A T T T C A A T T A C A A T T A C A A T
Db      5984 L y s g l i n a l g l i n a l e u l i e a p r o l i e s p g l i n a s n p r o a e n t h r a s p l y s g l u
Qy      406 -----
Db      6004 L y s g l i n a l e u l y s a p a r g l i e a s n g l i l e u e n g l i n g l y h i s a s n a p l e a s n
Qy      418 A A C G A A C C A T T A T T A T C T T A A A A G A A G C G G C A A C C A C T T G C C T T G A A
Db      6024 A s n a l a m e t t h r l y s g l u i n l i e l u g l i n a l a l y s a l a s g l i n e u l a g l i n a l e u g l i n
Qy      478 G A A C C A T T A T T A C C T C A A A G T T C A A A A G T T A A T T G C A A A A T T G A C C G A C T C A A T
Db      6044 A s p -----
Qy      535 -----
Db      6059 I l e l y s a l a l e u l a s n a l a l y s a r g a s p g l i n l i e a s n e r a s n p r o a p l e u t h r p r o
Qy      565 C A A G A C A C G G G C T T C C T T G A A A A G C A A T G A T T A A G T T A T G A T C T C A T A G A G G
Db      6079 G l u g l i n l y s a l a l y s a l a l e u l y s -----
Qy      625 C T T T C A A G A A T C A A G A A T T G C T C A A A A A T T G A C A A T C T C A A T C A A G G G A T C A
Db      6096 L e u -----

```

RESULT 5

```

Jc2299
Cell surface glycoprotein MSG99 - Pneumocystis carinii
C/Species: Pneumocystis carinii
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 03-Jul-2004
C/Accession: Jc2299
R/Model: M.; Nakamura, Y.
A/Title: MSG gene cluster encoding major cell surface glycoproteins of rat Pneumocystis
A/Reference number: Jc2299; MUID:96051989; PMID:8535973
A/Accession: Jc2299
A/Molecule type: DNA
A/Residues: 1-774 <WAD>
A/Cross-references: UNIPROT:Q01817; UNIPARC:UP100000694BC; GB:D31909; GB:D17441; NID:955
A/Gene: MSG99
C/Superfamily: Pneumocystis carinii major surface glycoprotein MSG100
C/Keywords: glycoprotein

```

Alignment Scores:

```

Pred. No.: 0.0822
Score: 127.50
Percent Similarity: 41.2%
Best Local Similarity: 22.4%
Query Match: 10.9%
Gaps: 2

```

US-09-360-685c-26 (1-685) x Jc2299 (1-774)

```

Qy      7 G G C A A A A T A G A T T T C A G A C A G T T A C G C A A A A A G C A C T T G A A A T T C C G T
Db      86 G l y s e u a p l e u v a l l y s e r l y s a l a l y s g l u l e c y a s a s n l l e g l u g l i n y s c y s
Qy      67 A A A G A T G A T C A T C A A A A G T A A C G A T A A G T T G A T A T C A T C A A C G G G T A
Db      106 L y s a s p -----
Qy      127 T C A G T G C T A A A G C A A C G G T G A T T C
Db      124 I l e i l e u a l a t h r l y s e r l i e p h e t y r g l u g l u c y a s e n l y s t y r p h e a s p t h r c y s
Qy      154 -----

```

```

Db      144 M e t h e p h e u l p r o t h r l y s a n g l i n g l i a l e a p h e c y s i l l e p h e a r s p
Qy      202 C A A T T G -----
Db      164 T h r c y s t r l e u i l e l y s g l i n l y s g l u m e t a r g a s n a l a l e u l y
Qy      241 G C T A G A -----
Db      184 A l a s p l l e t h r l y s e r l y s a p h e d i n l u l y s g l i n g l u l l e c y s e r t h r l e u
Qy      298 G T C G T A A T G G T T A T C T C A A C G A A C C A A C T T T G T A A A A C T T T C G A C A T C
Db      204 I l e l y s g l y a s p a p l e u l a s e r l e u c y s l e u g l u a e n l l e s e r p h e d l u a s p h e
Qy      358 A A G A A A G T T G A A T T G C A A A C T T G A A A T T T C A A T A C A A T A T A T G A C T C A A
Db      224 A s n a s a p l l e a p a r p h e c y s e l y s e r l e u s e r a s p l e u l l e a p a n g l y a r g l y s
Qy      418 A A C G A A C C A T T A T T A C T T A A T T A T A A A A A A A G A G G C A C A G T A G C T T G A A
Db      244 G l u g l i l l e c y s t r l u l y s l e u g l y -----
Qy      478 G A A C C A T T A C C T C A A A G T T C A A A A A G T A A T G C A A A A T T G A C C G A C T C A A T C A A
Db      253 -----
Qy      538 A T A G C A A G T G T T G C T T G T T G T A -----
Db      271 L e u c y s g l u d i l y s e l y r l e i l e y s e r p r o l l e u g l y p h e a n p r o l l e u g l u
Qy      589 A G C G A T A T A A A G T T A T C T C A A A A G T T A G T A G G T T
Db      291 A r g -----
Qy      631 A G A A A T C A A A A T T G 645
Db      309 A r g a s n a r g l y l e u 313

```

RESULT 6

```

D90011
Fnrb protein [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 03-Jul-2004
C/Accession: D90011
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu-
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shida, T.; Hattori, M.; Ogatawara, N.; Hayashi, H.; Hiramatsu, K.
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418186
A/Accession: D90011
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2481 <KDR>
A/Cross-references: UNIPROT:Q990R6; UNIPARC:UP1000000CAD87; GB:BA000018; PID:G13701961;
A/Experimental source: strain N315
A/Gene: fnbB(mrp)

```

Alignment Scores:

```

Pred. No.: 0.117
Score: 125.00
Percent Similarity: 40.9%
Best Local Similarity: 21.7%
Query Match: 10.7%
Gaps: 2

```

US-09-360-685c-26 (1-685) x D90011 (1-2481)

```

Qy      10 A A A A T A G A G A T T T C A G A A G T T A C G A C A A A A A G C A C T T G A A A T T C C G T T A A
Db      1374 G l u e n g l i n a l l i e a p a e n t h r t h r l y a l a t h r t h r g l u g l u y s a s n a l a l a l y s

```



```

Db      585 AsnSerSerLeuValSerIleSerGlnArgIleMetAspGluLeuAsnLysGlyGlnThr 604
Qy      409 GGAAGCTCAAAAGCAAGCCATTATGCTAAAGTTAAATAA-----AGAAAGCA 456
Db      605 AlaLeuSerSer-----ValGlnSerLysLeuSerThrIleAspGlnValIleAsnSer 622
Qy      457 GGGCAAGCAAGCTAGCCTTGAAGAACCCATTACGCTCAAGTGTCTAAAGTAATGCA 516
Db      623 GlyGlnIle-----IleLeuLysAsnGlyLysThr 632
Qy      517 AAAATTACCGAGCTCAATCAATAGCAAGGCTTTG-----GGTGT 558
Db      633 ArgIleAspArgLeuGlnThrValLeuProSerIleGlnGlnGlnIleSerAlaVal 652
Qy      559 GTAGGGCAAGCAGCGGGCTTCCCTTTGAAAGGATGATAAGTTGATGATTCAGTAAG 618
Db      653 LysAsnAlaGlnAlaAsnProLysValLysSerAspValAlaLysAlaAsnPhe 672
Qy      619 GTAGGGCTTCAAGATCAAA-----GAATGGCTCAGAAATTCGACATTCATCA 672
Db      673 Val-----ArgAsnAspLeuProGlnLeuGlnArgLeuThrAsnAlaThrAla 689
Qy      673 GCGGTATCAGAA 684
Db      690 SerValAsnLys 693

RESULT 10
E90598
membrane nuclease, lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C/Species: Mycoplasma pulmonis
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: E90598
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A>Title: The complete sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90598
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 11125 <KUR>
A:Cross-references: UNIPROT:Q98PM9; UNIPARC:UPI00000D464B; GB:AL445566; PID:g14090108; E
A:Experimental source: strain UAB CTIP
A:Genetics:
A:Gene: MYPV 6930
A:Genetic code: SGC3

Alignment Scores:
Pred. No.: 0.236 Length: 1125
Score: 121.00 Matches: 66
Percent Similarity: 39.1% Conservative: 41
Best Local Similarity: 24.1% Mismatches: 107
Query Match: 10.3% Indels: 60
DB: 2 Gaps: 13

US-09-360-685C-26 (1-685) x E90598 (1-1125)
Qy      4 AATGCGAAATATAGAT-----TTGAGCAAGGTATCGACGAAAGCAAGC 48
Db      581 AaAGlyGluAsnAspSerLysGlnAsnThrSerAsnSerArgGlnThrLysAsn 600
Qy      49 GACCTGAAATATCCGTTAAAGATGATGATCATCAATCAAAAGTAACGATTAAGTTGAT 108
Db      601 AspLeuArgSerLysGlnLysGlnAsnLeuThrThrLysAsnProSerSerAsnSer 620
Qy      109 AATCTC-----AATCAAGCGGTATCACTGCTAAAGCAAGC 144
Db      621 AsnValGluThrLysAsnGlnThrGlnAsnAsnGluAsnSerSerThrLysLysAspGlu 640
Qy      145 GGTGATTTCAATAGGAGTAAGCA-----GGCTTACCGCATCTCAAAAATTTCTCAAG 198
Db      641 IleAspThrSerLysThrGlnAspSerThrAsnSerAsnLeuLysAsnGlnLys 660
Qy      199 GAGCAATTGGCCCAACAGACTCAAAAATGAAGTCTCAATGCTAGA----- 246

```

```

Db      661 ThrAsnGlnValGlnThrLysThrAsnThrGluSerAsnSerSerThrAsnLys 680
Qy      247 -----AAAAATCTGAATATATCAATCCGTTAAGATGCTGTG 285
Db      681 GlnGlnGluAsnSerSerThrLysLysGlnGlnIleSerLysSerLysSerVal 699
Qy      286 AATGAAACCTTACTCGTATATGGCTTATCTCAAGCAAGAACCAACTCTTTCTAAAC 345
Db      700 AsnAsnSer-----AsnSerThrAsnLysGlnGln-----GluAsn 711
Qy      346 TTTTCGACATCAAGAAAGTTGAATGCAAACTTGAAATTTCAATAACAATCAAT 405
Db      712 IleAspAsnLysGlnLysGlnLysSerGluSerAsnValAsnAsnSerAsnSer 731
Qy      406 AATGACTCAAAACGACCCATTATGCTAAAGTTAATAAAG----- 450
Db      732 ThrAsnThrGlnAsnGlnGlnThr---ProGluThrAsnGlnSerGlnAsnValIle 750
Qy      451 -----AAGCAAGGCAAGCAGCTAGCTTGAAGAACCATTTACGCTCAAGTGTAA 504
Db      751 IleGlyLysAsnProAsnAsnGlnSerLeuAsnAlaIleAspValSerAlaLys 770
Qy      505 AAGTAAATGCAAAATATGACCGACTCAATCAATAGCAAGTGTGCTGTGTAGGG 564
Db      771 LysVal-----LysIleGlyTyrTrpAsn-----IleAsnGluSerValGlyLysSer 787
Qy      565 CAGCAGCGGGCTTCCCTTTGAAAG-----CATGATTAAGTTGATGATCTCACT 615
Db      788 AlaSerLysAlaPheLeuValAlaLysValIleAspHisAsnLysLeuAspLeuValGly 807
Qy      616 AAGTGGGCTTTCAGAGATCAAGATGATGCTCAAAAATTCGACATCTCAAT----- 669
Db      808 IleGlyLysLeuValHisGlnGlnThrLeuThrLysLysValGlnGlnLysLeu 827
Qy      670 -----CAGCGGTATCAGAA 684
Db      828 SerLysAspSerSerAspLysTrpValGlnValIleSerGln 841

```

```

RESULT 11
T28676
rhodopy protein - Plasmodium yoelii (fragment)
C/Species: Plasmodium yoelii
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T28676; A45521
R:Simha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A>Title: Comparison of two members of a multigene family coding for high-molecular mas.
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784;
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhodopy protein. Multiple
A:Accession: A45521; MUID:91101660; PMID:2270106
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEB>
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Alignment Scores:
Pred. No.: 0.228 Length: 2401
Score: 121.00 Matches: 50
Percent Similarity: 37.9% Conservative: 47
Best Local Similarity: 19.5% Mismatches: 87
Query Match: 10.3% Indels: 72
DB: 2 Gaps: 9

US-09-360-685C-26 (1-685) x T28676 (1-2401)

```


erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2641
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <KUR>
A:Cross-references: UNIPROT:O8UHV2; UNIPARC:UPI00001644D8; GB:AE008688; PIDD:AA141544.1
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: mcIA
A:Map position: circular chromosome
C:Superfamily: Methyl-accepting chemotaxis protein mcra

Alignment Scores:

Pred. No.:	0.339	Length:	615
Score:	119.00	Matches:	52
Percent Similarity:	42.0%	Conservative:	45
Best Local Similarity:	22.5%	Mismatches:	94
Query Match:	10.2%	Indels:	40
DB:	2	Gaps:	9

US-09-360-685C-26 (1-685) x AB2641 (1-615)

OY 88 AAGGTAAACGATTAAGTGTATATCTCAATCAAGC-----GTATCACTG 132

Db 226 LysLeuThrAspAlaValLysSerLeuSerAspGlyAspLeuGluThrProIleProTyr 245

OY 133 GCTAAAGCAGCGGATTCAGTAGAGCGTAAAGACAGACGATTCGAAATTTTC 192

Db 246 AlaThrAsnThrAsnGluPheGlyAsnIleAlaArgAlaLeuValIlePheArgGluAsn 265

OY 193 TCAAGAGCGCAITGGC-----CAACAGCTCAAAA 225

Db 266 AlaIleGluLysLeuAlaIleGluGlyLysSerAlaGluArgSerAlaIleGluSer 285

OY 226 AATGAATCTCAATGCTAGAAAATAATCTGAATATATTCATCCGTTAAGATGCTGTG 285

Db 286 GluArgHisArgAsnAspAlaGluIleGluLeuAspGlyGlnIleGluPheAlaVal 305

OY 286 AAT-----GGAACCTTACTCGGTATGGGTATCTCAAGACAGAACCC 327

Db 306 GlyIuIleAlaSerGlyLeuGlyArgLeuSerArgGlyAspLeuSerArgThrIleGlu 325

OY 328 ACAACTTTCTTAAAACTTTTCGACATCAAGAAAGAGTTGAATGCAAACTTGAAT 387

Db 326 ThrProPheAlaGlyArgLeuAspArgLeuArgThrAspPheAsnGluSerLeuLeuAsn 345

OY 388 TTCATTAACAT-----AACATTAATGCACTC 414

Db 346 LeuArgAspAlaLeuGlyGlnIleArgGluArgThrLeuIleIleGlnAsnSerGlyIle 365

OY 415 AAAAGACGACCCATTATGCTAAAGTTAATTAATAAGAAAGACAGGCGAAGAGCTT 474

Db 366 GluIleGluGlnSerSerValAspLeuSerLysArgThrGluSerGlnAlaAlaSerLeu 385

OY 475 GAAGAACCCATTATGCT-----CAAGTTGCTAAAAAGTTAATGCAAAATTTGACCGA 528

Db 386 GluGluThrAlaAlaAlaValaGluGluIlePheAlaThrValaArgSerSerAlaGluArg 405

OY 529 CTCATCAAAATAGACAGTGTGGGTGTTGAGGCGACAGCAGC-----GGCTTC 579

Db 406 AlaArgGluAlaAsnGluAlaValaArgValThrLysGlnSerAlaAspSerSerGlySer 425

OY 580 CCTTGAAGAAAGCATGATTAAGTGTAGATCTCAGTAAGGA---GGGCTTCAAGGAAT 636

Db 426 ValValSer-----AsnAlaValaAspAlaMetSerLysIleGluGlyAlaSerGlyVal 443

637 ---CAAGAAATTGGCTCAGAAAATTGACATCTC 666

Db 444 l1egluGln1le1legluVal1leAspAsp1le 454

RESULT 14

D97423

mcia protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 05-Oct-2004
C:Accession: D97423

R:Goodner, B.; Hinkle, G.; Gettung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2333-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium t
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D97423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <KUR>
C:Cross-references: UNIPROT:Q8UHY2; UNIPARC:UPI00000D83A; GB:AE007869; PIDN:AAK6341.

A:Gene: AGR_C_929
C:Map position: circular chromosome
C:Superfamily: Methyl-accepting chemotaxis protein mcpA

Alignment Scores:

Pred. No.:	0_339	Length:	622
Score:	119.00	Matches:	52
Percent Similarity:	42.0%	Conservative:	45
Best Local Similarity:	22.5%	Mismatches:	94
Query Match:	10.2%	Indels:	40
DB:	2	Gaps:	9

US-09-360-685C-26 (1-685) x D97423 (1-622)

QY 88 AAGTAAACGATPAAAGTATATCTCAATCAACG-----GTATCAGTG 132

Db 233 lylsleuThraAspAlaValylsSerleuSerAspGlyAspleuGlnThrProileProtyr 252

QY 133 GGTAAAGCAACGGGTGATTTCAGTAGGGTGAAGCAACGCTTACCCGATCTCAAAAATTTC 192

Db 253 AlathraenThraSngluPhelGluYasn1leAlaArgAlaLeuVal1lePheArgGluAsn 272

QY 193 TCAAAAGAGCAATTTGGCC-----CAACAAGCTCAAAA 225

Db 273 Ala1leGluYsleuAla1leGluGlylsSerAlaGlnGluArgSerAla1leGluSer 292

QY 226 AATGAAGCTCTCAATGCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGGTGTG 285

Db 293 GluArgH1sArgAsnaAspAlaGluYsGlnGluLeuAspGlyGln1leGluPhelAlaVal 312

QY 286 AAT-----GGAACCTTAGTCGGTAAATGGGTTATCTCAAGCAAGACC 327

Db 313 GlnGlu1leAlaSerGlyLeuGlnYArgLeuSerArgGlyAspLeuSerArgThrlleGlu 332

QY 328 ACAACTCTTTTAAAAAATTTTCGAGCATCAAGAAAGAGTGAATGCAAAATCTGGAAT 387

Db 333 ThrProPheAlaGlyArgLeuAspArgLeuArgThraSPheAsnGlnUserLeuLeuAsn 352

QY 388 TTCATTAACAAT-----AACATAATGAGACTC 414

Db 353 LeuArgAspAlaLeuGlnYGln1leArgGluArgThrlleU1le1leGlnHensArgY1le 372

QY 415 AAAAAAGAACCCATTATGCTAAAGTTAATAAAAAGAAAGCAGGCGAAGCGTACGCTT 474

Db 373 Gln1leGlnGlnSerSerValAspLeuSerLysArgThrgluAsnGln1leAla1leSerLeu 392

QY 475 GAAGAACCATTATAGCT-----CAAGTGTCTAAAAAGCTTAATCAAAATATGACCGA 528

Db 393 GlnGlnThraAlaAlaValaGlnGlnU1leThra1leThraValArgSerSerAlaGluArg 412

QY 529 CTCAATCAAAATAGCAAGTGGTGTGGGTGTGTGAAGGCAAGCAGCG-----GGCTTC 579

Db 413 Ala1leArgGluAlaAsnGlnAlaValaArgValaThrlYsGlnSerLysAsnSerSerArgY1Ser 432

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 20:17:32 ; Search time 2.8 Seconds
(without alignments)
978.592 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcgaataaaga.....caatcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 135346 segs, 20000420 residues

Total number of hits satisfying chosen parameters: 270692

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.n2p.model -DEV=klp
-O=/abs/ABSGWES.spool/US09360685/runat.06032006.121329.15688/app.query.fasta.1
-DB=Published Applications AA New -OPMT=faatan -SUFFIX=rapbn -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-USER=US09360685@CGN1_1_17@runat.06032006.121329.15688 -NCPU=6 -ICPU=3
-NO KWAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdp.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdp.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdp.*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdp.*
5: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pdp.*
6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdp.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pdp.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdp.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	6	US-10-615-668-5
2	896.5	76.6	1167	7	US-11-052-554A-121
3	127	10.8	1290	6	US-10-485-517-141
4	125	10.7	1448	6	US-10-485-517-212
5	120.5	10.3	5024	6	US-10-793-626-2364
6	118	10.1	1713	6	US-10-766-317-2
7	118	10.1	1724	6	US-10-766-317-6
8	118	10.1	3333	6	US-10-766-317-4
9	117	10.0	693	7	US-11-196-475-68

10	115.5	9.9	1095	6	US-10-793-626-3154	Sequence 3154, App
11	114	9.7	3712	7	US-11-019-711-48	Sequence 48, Appl
12	112.5	9.6	611	6	US-10-793-626-2586	Sequence 2586, Ap
13	112	9.6	3712	7	US-11-019-711-51	Sequence 51, Appl
14	111.5	9.5	700	7	US-11-196-475-66	Sequence 66, Appl
15	110.5	9.4	708	7	US-11-196-475-76	Sequence 76, Appl
16	110.5	9.4	739	7	US-11-087-099-12273	Sequence 12273, A
17	109.5	9.4	2087	7	US-11-075-185-285	Sequence 28, Appl
18	109	9.3	2480	6	US-10-995-561-825	Sequence 825, App
19	109	9.3	3116	6	US-10-995-561-826	Sequence 826, App
20	108	9.2	401	7	US-11-097-749-3	Sequence 3, Appl
21	108	9.2	863	7	US-11-097-749-2	Sequence 74, Appl
22	105.5	9.0	700	7	US-11-196-475-74	Sequence 169, App
23	105.5	9.0	1404	6	US-10-878-556A-169	Sequence 169, App
24	104.5	8.9	1652	6	US-10-995-561-663	Sequence 663, App
25	104.5	8.9	1938	6	US-10-995-561-661	Sequence 661, App
26	104.5	8.9	1938	6	US-10-995-561-662	Sequence 662, App
27	104.5	8.9	1954	6	US-10-995-561-660	Sequence 660, App
28	104.5	8.9	1972	6	US-10-995-561-664	Sequence 664, App
29	104.5	8.9	1972	6	US-10-995-561-666	Sequence 666, App
30	104	8.9	2107	6	US-10-995-561-627	Sequence 827, App
31	103	8.8	440	6	US-10-976-933-2	Sequence 2, Appl
32	102.5	8.8	1279	6	US-10-793-626-3188	Sequence 3188, Ap
33	102.5	8.8	663	7	US-11-196-475-70	Sequence 70, Appl
34	102	8.7	663	7	US-11-196-475-78	Sequence 78, Appl
35	102	8.7	663	7	US-11-196-475-78	Sequence 78, Appl
36	101.5	8.7	380	7	US-11-144-833-13	Sequence 13, Appl
37	101	8.6	2036	7	US-11-124-368A-276	Sequence 276, App
38	101	8.6	2036	7	US-11-124-368A-280	Sequence 280, App
39	101	8.6	2036	7	US-11-124-368A-281	Sequence 281, App
40	101	8.6	2044	7	US-11-124-368A-278	Sequence 278, App
41	101	8.6	2144	6	US-11-124-368A-277	Sequence 277, App
42	98.5	8.5	794	6	US-10-793-626-1050	Sequence 1050, Ap
43	98.5	8.4	1189	7	US-10-689-742-142	Sequence 142, App
44	98.5	8.4	1189	7	US-11-074-176-134	Sequence 134, App
45	98	8.4	239	6	US-10-485-517-249	Sequence 249, App

ALIGNMENTS

RESULT 1
US-10-615-668-5
; Sequence 5, Application US/10615668
; Publication No. US20050276819A1
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides
; FILE REFERENCE: CHIR0337
; CURRENT APPLICATION NUMBER: US/10/615,668
; PRIOR FILING DATE: 2003-07-08
; PRIOR APPLICATION NUMBER: 08/471,491
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/256,848
; PRIOR FILING DATE: 1994-10-21
; PRIOR APPLICATION NUMBER: 09/410,835
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-615-668-5
Alignment Scores:
Pred. No.: 5,27e-76
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Length: 1147
Matches: 228
Conservative: 0
Mismatch: 0

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 2

Query Match: 96.2% Indels: 0
DB: 6 Gaps: 0
US-09-360-685c-26 (1-685) x US-10-615-668-5 (1-1167)

```
QY 1 AAAAATGCGAATAAAGATTTCAGCAAGGTAAGCAAGCAAAAGCACTTGAAAT 60
DB 750 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGTCAATCA 120
DB 770 SerValLysaenValLysaenValLysaenValLysaenValLysaenValLysaen 789
QY 121 GCGGTATCGTGTAAAGCAAGGTAAGTTCAGTAAGGTAAGCAAGCAAGTTCAG 180
DB 790 AlaValSerValAlaLysaenLysaenLysaenLysaenLysaenLysaenLysaen 809
QY 181 CTCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGTCAAT 240
DB 810 LeuLysaenPheSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 829
QY 241 GGTGAAAAAATCTGAATATATCAATCCGTTAAGAAATGTTGTAAGCAAGTCA 300
DB 830 AlaLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 849
QY 301 GGTAAATGTTATCTCAAGCAAGCAAGCAAGTTCGTTAAAGTTCGTTGCAATCA 360
DB 850 GlyaenGIyLysaenGIyLysaenGIyLysaenGIyLysaenGIyLysaenGIyLysaen 869
QY 361 AAAGAGTTGAATGCAAACTTGAAATTCATTAACAATTAACAATTAACAATTAACA 420
DB 870 LysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 889
QY 421 GAACCAATTTATGTTAAAGTAAATTAAGAAAGCAAGCAAGCAAGTTCGTTGAA 480
DB 890 GluProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 909
QY 481 CCCATTTAGCTCAAGTTCGTTAAAGTAAATTAAGCAAAATTAAGCAATCAATCA 540
DB 910 ProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 929
QY 541 GCAAGTGTGGTGTGTTGTAAGGCAAGCAAGCAAGTTCGTTGAAAGCAATCAATCA 600
DB 930 AlaSerGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 949
QY 601 GTTATGATCTCAGTAAGGTAAGGTTTCAGCAAGTCAAGAAATGCTGCAAGAAATGAC 660
DB 950 ValaPhePheLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 969
QY 661 AATTCATCAAGGCAATCAAG 684
DB 970 AenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 977
```

RESULT 2

US-11-052-554A-121
; Sequence 121, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-05-11/052,554A
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent version 3.3
; SEQ ID NO 121
; LENGTH: 1167
; TYPE: PRT

i ORGANISM: Helicobacter pylori J99
US-11-052-554A-121

Alignment Scores:

Pred. No.:	4,496-59	Length:	1167
Score:	896.50	Matches:	189
Percent Similarity:	74.8%	Conservative:	7
Best Local Similarity:	72.1%	Mismatches:	15
Query Match:	76.6%	Indels:	51
DB:	7	Gaps:	2

US-09-360-685c-26 (1-685) x US-11-052-554A-121 (1-1167)

```
QY 1 AAAAATGCGAATAAAGATTTCAGCAAGGTAAGCAAGCAAAAGCACTTGAAAT 60
DB 754 LysaenGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 773
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGTCAATCA 120
DB 774 SerValLysaenValLysaenValLysaenValLysaenValLysaenValLysaen 793
QY 121 GCGGTATCGTGTAAAGCAAGGTAAGTTCAGTAAGGTAAGCAAGCAAGTTCAG 180
DB 794 AlaValSerValAlaLysaenLysaenLysaenLysaenLysaenLysaenLysaen 813
QY 181 CTCGTTAAAGATGATCATCAATCAAAAGGTAAGGTAAGGTAAGTCAAT 240
DB 814 LeuLysaenPheSerLysaenLysaenLysaenLysaenLysaenLysaenLysaen 833
QY 241 GGTGAAAAAATCTGAATATATCAATCCGTTAAGAAATGTTGTAAGCAAGTCA 300
DB 834 ThrGIyLysaenSerAlaLysaenLysaenLysaenLysaenLysaenLysaenLysaen 853
QY 301 GGTAAATGTTATCTCAAGCAAGCAAGCAAGTTCGTTAAAGTTCGTTGCAATCA 360
DB 854 GlyaenGIyLysaenGIyLysaenGIyLysaenGIyLysaenGIyLysaenGIyLysaen 873
QY 361 AAAGAGTTGAATGCAAACTTGAAATTCATTAACAATTAACAATTAACAATTAACA 420
DB 874 LysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 893
QY 421 GAACCAATTTATGTTAAAGTAAATTAAGAAAGCAAGCAAGCAAGTTCGTTGAA 480
DB 894 Ser-----ThrGlu 896
QY 481 CCCATTTAGCTCAAGTTCGTTAAAGTAAATTAAGCAAAATTAAGCAATCAATCA 540
DB 897 ProLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 916
QY 541 GCAAGTGTGGTGTGTTGTAAGGCAAGCAAGCAAGTTCGTTGAAAGCAATCAATCA 600
DB 917 AlaSerGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 936
QY 576 ----- 576
DB 937 ValaPhePheLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 956
QY 577 -----TTCCGTTGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 956
DB 957 AaPhePheLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 976
QY 619 GTAGGCTTCAGCAAGTTCAGCAAGTTCAGCAAGTTCAGCAAGTTCAGCAAGTTCAG 678
DB 977 ValGIyLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaenLysaen 996
QY 679 TCAGAA 684
DB 997 SerGlu 998
```

RESULT 3

US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256239A1

```

; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 1290
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-485-517-141

Alignment Scores:
Pred. No.: 0.0148 Length: 1290
Score: 127.00 Matches: 69
Percent Similarity: 37.1% Conservative: 32
Best Local Similarity: 25.4% Mismatches: 93
Query Match: 10.8% Indels: 78
DB: 6 Gaps: 12

US-09-360-685C-26 (1-685) x US-10-485-517-141 (1-1290)
QY 7 GCGAATAATAGAGATTTCAGACAGGTAACGCAAGCAAAAGGACCTTGAATAATTCCTT 66
DB 181 G1yAenGluGlnIlyeAlaAsnValAspAlaAlaMetAspLysLeuArgGlnSerIle 200
QY 67 AAAGATGTCATC-----AATCAAAAGTAAACGAT-----AAAGTT 105
DB 201 AlaAspAsnAlaThrThrysglnAsnGlnAsnYrThAspAlaSerGlnAsnIlySls 220
QY 106 GATTAATCTCAATCAAGCGGTATCAGTGGCT----- 135
DB 221 AspAlaTyAsnAsnAlaValThrThrylGlnGlnIlyleIleAspGlnThrThrySerPro 240
QY 136 -----AAAGCAAGCGGTGATTTGATGAGGTAGACGACG 171
DB 241 ThrLeuAspProThrValIleAsnGlnAlaAlaGlnValAspThrThryAsnAla 260
QY 172 TTAGCGCATCTCAAAATTTTC-----TCAAGAGCAATTTGGCCCAACGCTCAAAA 225
DB 261 LeuAsnGlnValAsnGlnIlyAsnGlnIlyAlaIlyGlnGlnIlyAlaSerGlnSerLeuGlySer 280
QY 226 AATGAAAGTCTGATGCTGAGAAAATAATGTAATATATCAATCCGTTAAGATGCTG 285
DB 281 LeuAspAsnLeuAsnAlaGlnIlyS-----GlnThrValThrAspGlnIle 296
QY 286 AATGGA----- 291
DB 297 AsnGlnAlaIlyThrValAspGlnAlaAsnGlnIlyeGlnAsnAlaGlnAsnLeuAsn 316
QY 292 ACCGTAGTGGTAAAT-----GGGTATCTCAAGCAAGCAACCTTTCTAAA 342
DB 317 ThrAlaMetGlyAsnLeuIlyGlnAlaIleAlaAspIlyAspAlaThrIlyeAlaThrVal 336
QY 343 AACTTTTGGACATCAAGAAAGATGGAATGCAAACTTGAAATTCATTAACAATAC 402
DB 337 AsnPhetThrAspAlaAspGln-----AlaTyGlnGlnAlaIlyAsnThrAlaVal 353
QY 403 AATATGAGCTGAAAAAGAACCACTTATGCTAAAGTAAATTAATAAGAAAGACGCGCA 462
DB 354 ThrAsnAlaI-----GlnAsnIleIleSerIlyValAsnGlnIlyGlnAlaThrGln 370
QY 463 GCAAGTAGCTTGAAGAACCACTTACGCTCAAGTTGCTAAAGAGTAAATGCAAAATTT 522

```

```

DB 371 AlaGlu-----ValGlnAlaIleIlyGlnValAsnAlaIlySls 384
QY 523 GACGACATCAATCAATAGCAAGTGGTTGGGTGTGGAGGCAAGACGGCTTCCT 582
DB 385 GlnAlaLeuAsnGlyAsnAlaAsn----- 392
QY 583 TTGAAAGCATGATTAAGTATGATGATCTAGTAAGTAGGGCTTTCAAGATCAAGAA 642
DB 393 ---ValGlnIlyAlaTyAspGlnAlaThrAlaLeuIleAsnSerAsnAspLeuAsn 411
QY 643 TTGGCTCAAAATTCACAAATCTCAATCAAGCGGTA 678
DB 412 GlnAlaGlnIlyS-----AspAlaLeuIlyGlnGlnVal 422

RESULT 4
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-485-517-212

Alignment Scores:
Pred. No.: 0.0207 Length: 1448
Score: 125.00 Matches: 51
Percent Similarity: 40.9% Conservative: 45
Best Local Similarity: 21.7% Mismatches: 77
Query Match: 10.7% Indels: 62
DB: 6 Gaps: 8

US-09-360-685C-26 (1-685) x US-10-485-517-212 (1-1448)
QY 10 AAAAATAAGATTTCAGCAAGGTAAACGCAAGCAAAAGCCTTGAATAATTCCTTAAA 69
DB 361 GlnAsnGlnAlaIleAspAsnThrThrGlyAlaThrThryGlnGlnIlyAsnAlaIlySls 380
QY 70 GATGTGATCATCAATCAAAAGGTAAAGGTAAAGTATATTCATCAAGCGGTATCA 129
DB 381 AspLeuValLeuIlyValAlaTyGlnIlyAlaTyGlnAspIleLeuAsn----- 396
QY 130 GTGGCTAAGCAAGCGGTGATTCAGTAGGGTA-----GAGCAAGCTTAGGCCGAT 180
DB 397 ---AlaGlnThrThrAsnAspValThrGlnIlyeAspGlnAlaValAlaAspIlyeGln 415
QY 181 -----CTCAAAATTTCTCAAAAGCAAGCAATTTGCGTTAAGAT 219
DB 416 GlyIleThrAlaAspThrThrIlyeAspValAlaIlyAspGlnAlaIlyAspIlyeAla 435
QY 220 CAATAAATAAGATCTCAATGCTGAGAAAATAATGTAATATATCAATCCGTTAAGAT 279
DB 436 AsnGlnIlyIlyValAlaLeuIleAlaGln----- 444
QY 280 GGTGTGAATGGAACCTTAGTGGTATGCGTTATCTCAAGCAAGCAACCTTTCT 339
DB 445 -----ThrAlaAspAlaThrThrGlnGln 452

```


Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 4

```
OY 340 AAAAATTCTTGGACATCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 399
DB 453 Lys-----GluGlnAlaIleuGlnIleValAlaIleuThrGlnIleValAlaIleu 470
OY 400 AACATATATGAGCTCAAAAAGCAACCATTTTGTCTAAAGTTATTAAGAAAGCAAG 459
DB 471 -----IleGlnAlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleu 485
OY 460 CAAGCAGCTAGCCTTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 519
DB 486 AsnAlaIleuGlnAlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleu 505
OY 520 ATGACGAGCTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 579
DB 506 AlaGluIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThrGlnIleVal 512
OY 580 CTTTGAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 639
DB 513 -----GlnAsnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleu 528
OY 640 GAATGAGCTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 684
DB 529 GluIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThrGlnIleVal 543
```

```
RESULT 5
US-10-793-626-2964
/ Sequence 2964, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMBERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PUS480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ PRIOR FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: Patent Ver. 2.1
/ SEQ ID NO 2964
/ LENGTH: 5024
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
/ NAME/KEY: MOD_RES
/ LOCATION: (5024)
/ OTHER INFORMATION: variable amino acid
US-10-793-626-2964
```

```
Alignment Scores:
Pred. No.: 0.0428 Length: 5024
Score: 120.50 Matches: 62
Percent Similarity: 41.9% Conservative: 50
Best Local Similarity: 23.2% Mismatches: 84
Query Match: 10.3% Indels: 71
DB: 6 Gaps: 13
```

US-09-360-685c-26 (1-685) x US-10-793-626-2964 (1-5024)

```
OY 28 AAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 84
DB 74 LysLeuThrAlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 93
OY 85 CAAGCAGCTAGCCTTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 144
DB 94 -----ThrProGlnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleu 110
OY 145 GGTGATTCAGTGGTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 204
DB 111 ThrAspValThr-----AlaGluIleuThrGlnIleValAlaIleuThrGlnIleVal 129
```

```
OY 205 TTGGCCCAACAGCTCAAAAAGTGTGATTCGAAAATTGGAATTCATACAT 243
DB 130 LeuThrGlnAlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 149
OY 244 AGAAAAA-----TCTGAATATATCAATTCGAAAATTGGAATTCATACAT 288
DB 150 GluProGlnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 169
OY 289 -----GGAACCTTGAAGTGTGATTCGAAAATTGGAATTCATACAT 336
DB 170 GluGlnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 189
OY 337 TCTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 375
DB 190 ThrIleAsnAlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 209
OY 376 -----AACTTGAAGTGTGATTCGAAAATTGGAATTCATACAT 417
DB 210 GluIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThrGlnIleVal 229
OY 418 AACGAACCATTTATGCTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 477
DB 230 AsnSer-----SerGlnIleuThrGlnIleValAlaIleuThrGlnIleVal 244
OY 478 GAACCATTTATGCTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 537
DB 245 -----AlaIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 579
OY 538 ATAGCAAGTGTGATTCGAAAATTGGAATTCATACAT 579
DB 258 LeuIleAsnGlyIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 599
OY 580 -----CTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 600
DB 278 GlnGlnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThrGlnIleVal 297
OY 601 GTTGTATGCTCAAGAAAGTGTGATTCGAAAATTGGAATTCATACAT 660
DB 298 AsnProGlnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThr 311
OY 661 AATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 681
DB 312 AsnIleuThrGlnIleValAlaIleuThrGlnIleValAlaIleuThrGlnIleVal 318
```

```
RESULT 6
US-10-766-317-2
/ Sequence 2, Application US/10766317
/ Publication No. US20060009630A1
/ GENERAL INFORMATION:
/ APPLICANT: Markovach, M. Peter
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
/ FILE REFERENCE: 33828/US/R/PT/R/S
/ CURRENT APPLICATION NUMBER: US/10/766,317
/ PRIOR FILING DATE: 2004-01-27
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Patent version 3.2
/ SEQ ID NO 2
/ LENGTH: 1713
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-766-317-2
```

```
Alignment Scores:
Pred. No.: 0.0674 Length: 1713
Score: 118.00 Matches: 55
Percent Similarity: 39.4% Conservative: 407
Best Local Similarity: 21.2% Mismatches: 50
Query Match: 10.1% Indels: 9
DB: 6 Gaps: 9
```

US-09-360-685c-26 (1-685) x US-10-766-317-2 (1-1713)

OY 13 AATAAGATTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 72

```

Db      300 AsnAsnValAsnArgAlaThrGlnSerAlaLysGlnLeuAspValLysIleLysAsn 319
Qy      73 GTGATCATCAATCAAAAGGTACCGATTAAGTTGATATCTCAATCAAGCGGTATCACTG 132
Db      320 ValIleArgAsnValHisIleLeuLeuLysGlnIleSerGlyThrAspGlyGlnLysAsn 339
Qy      133 GCTAAAGCAACGGGTGATTTCACTAGTACG-----GTAGAGCAAGCGTTAGCCGAT 180
Db      340 AsnValProSerGlyAspPheSerArgGlnTrpAlaGlnLacGlnMetCysArgGln 359
Qy      181 CTC-----AAAAATTTCTCAAAAGAGCAATTGGCCCAACAGCTCAAAAATAAGT 234
Db      360 LeuArgAsnArgAsnPheGlyLysHisLeuArgGlnLacGlnLacAspLysArgGlnSer 379
Qy      235 CTCAATGCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGGTGTGAATGGAACC 294
Db      380 GlnLeuLeuLeuAsnArgIleArgGlyThrGlnLysThrHisGlnGlnLysAsn----- 397
Qy      295 CTAGTCGGTAAATGGTTATCTCAAGCAAGCAACCAACTCTTCTAA-----AAC 345
Db      398 -----AsnGlyLeuAlaAsnSerIleArgAspSerLeuAsnGlnLysGlnLys 414
Qy      346 TTTTGGACATCAAGAAAGAGTGAATGCAAAATTGAAATTTCAATACATTAACAAT 405
Db      415 LeuSerAspLeuArgAlaArgLeuGlnLysAlaLacGlnLacLysGlnLysGlnLys 434
Qy      406 AATGACTCAAAAACCAACCATTTATGCTAAAGTTAATAAAG----- 450
Db      435 LeuAsnGlnLysAsnGlnLysArgAlaLeuGlnLysAlaLacGlnLysGlnLys 454
Qy      451 -----AAGCAGGGCAAGCACTAGCTTGAAGA 480
Db      455 SerLeuGlnSerAspPheThrLysTrpLeuThrThrAlaAspSerLeuLeuGlnThr 474
Qy      481 CCCATTATACGCTCAAGTTGCTAAAGGTAAATGCAAAATGACCGACTCAATCAATA 540
Db      475 AsnIleAlaLeuGlnLeuMetGlnLysSerGlnLysGlnLysLeu----- 491
Qy      541 GCAAGTGGTTGGGTGTTGTAGGGCAAGCAGCGGCTTCCCTTGAAGAGCATGATAA 600
Db      492 AlaAlaSerLeuAsnGlnLysAlaArgGln-----GlnLeuSerAspLys 505
Qy      601 GTTGATGATCTCAGT-----AAGTAGGGCTTCAAGAAAT----- 636
Db      506 ValArgGlnLeuSerArgSerAlaGlyLysThrSerLeuValGlnLysAlaGlnLysHis 525
Qy      637 -----CAAGATTTGGCTCAGAAAATTGACATCTCAATCAAGCGGTATCA 681
Db      526 AlaArgSerLeuGlnGlnLysAlaLysGlnLeuGlnGlnLysAlaArgAsnAlaSer 544

RESULT 7
US-10-766-317-6
; Sequence 6, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1724
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-6

Alignment Scores:
Pred: 0.0674 Length: 1724
Score: 118.00 Matches: 55
Percent Similarity: 39.4% Conservative: 47

```

```

Best Local Similarity: 21.2% Mismatches: 107
Query Match: 10.1% Indels: 50
DB: 6 Gaps: 9

US-09-360-685C-26 (1-685) x US-10-766-317-6 (1-1724)

Qy      13 AATTAAGATTTAGAGCAAGGTAAACGCAAAAAGCACTTGAATAATTCGTTAAGAT 72
Db      311 AsnAsnValAsnArgAlaThrGlnSerAlaLysGlnLeuAspValLysIleLysAsn 330
Qy      73 GTGATCATCAATCAAAAGGTACCGATTAAGTTGATATCTCAATCAAGCGGTATCACTG 132
Db      331 ValIleArgAsnValHisIleLeuLeuLysGlnIleSerGlyThrAspGlyGlnLysAsn 350
Qy      133 GCTAAAGCAACGGGTGATTTCACTAGTACG-----GTAGAGCAAGCGTTAGCCGAT 180
Db      351 AsnValProSerGlyAspPheSerArgGlnTrpAlaGlnLacGlnMetCysArgGln 370
Qy      181 CTC-----AAAAATTTCTCAAAAGAGCAATTGGCCCAACAGCTCAAAAATAAGT 234
Db      371 LeuArgAsnArgAsnPheGlyLysHisLeuArgGlnLacGlnLacAspLysArgGlnSer 390
Qy      235 CTCAATGCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGGTGTGAATGGAACC 294
Db      391 GlnLeuLeuLeuAsnArgIleArgGlyThrGlnLysThrHisGlnGlnLysAsn----- 408
Qy      295 CTAGTCGGTAAATGGTTATCTCAAGCAAGCAACCAACTCTTCTAA-----AAC 345
Db      409 -----AsnGlyLeuAlaAsnSerIleArgAspSerLeuAsnGlnLysGlnLys 425
Qy      346 TTTTGGACATCAAGAAAGAGTGAATGCAAAATTGAAATTTCAATACATTAACAAT 405
Db      426 LeuSerAspLeuArgAlaArgLeuGlnLysAlaLacGlnLysGlnLysGlnLys 445
Qy      406 AATGACTCAAAAACCAACCATTTATGCTAAAGTTAATAAAG----- 450
Db      446 LeuAsnGlnLysAsnGlnLysArgAlaLeuGlnLysAlaLacGlnLysGlnLys 465
Qy      451 -----AAGCAGGGCAAGCACTAGCTTGAAGA 480
Db      466 SerLeuGlnSerAspPheThrLysTrpLeuThrThrAlaAspSerLeuLeuGlnThr 485
Qy      481 CCCATTATACGCTCAAGTTGCTAAAGGTAAATGCAAAATGACCGACTCAATCAATA 540
Db      486 AsnIleAlaLeuGlnLeuMetGlnLysSerGlnLysGlnLysLeu----- 502
Qy      541 GCAAGTGGTTGGGTGTTGTAGGGCAAGCAGCGGCTTCCCTTGAAGAGCATGATAA 600
Db      503 AlaAlaSerLeuAsnGlnLysAlaArgGln-----GlnLeuSerAspLys 516
Qy      601 GTTGATGATCTCAGT-----AAGTAGGGCTTCAAGAAAT----- 636
Db      517 ValArgGlnLeuSerArgSerAlaGlyLysThrSerLeuValGlnLysAlaGlnLysHis 536
Qy      637 -----CAAGATTTGGCTCAGAAAATTGACATCTCAATCAAGCGGTATCA 681
Db      537 AlaArgSerLeuGlnGlnLysAlaLysGlnLeuGlnGlnLysAlaArgAsnAlaSer 555

RESULT 8
US-10-766-317-4
; Sequence 4, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 3333
; TYPE: PRT

```

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbn

Page 6

ORGANISM: Homo sapiens
US-10-766-317-4

Alignment Scores:

Pred. No.: 0.0661 Length: 3333
Score: 118.60 Matches: 55
Percent Similarity: 39.48 Conservative: 47
Best Local Similarity: 21.24 Mismatches: 107
Query Match: 10.14 Indels: 50
Gaps: 9

US-09-360-685c-26 (1-685) x US-10-766-317-4 (1-3333)

```
QY 13 AATAAGCATTTGACGACGTAACGCAAGCAAAAGGACCTTGAAATTCGGTTAAGAT 72
DB 1920 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 73 GTGATCATCATCAAAAGGTAACGATTAAGTGAATTCATCAATCAACGCGTCACTG 132
DB 1940 VALLIATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
QY 133 GCTAAGCAACGCGTATTCATGAGG-----GTAGACGACGTTAGCCGAT 180
DB 1960 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 181 CTC-----AAAATTTCTCAAGAGCAATTCGCGCCACGACGTTCAAAATGAAAGT 234
DB 1980 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 235 CTCAATGCTAAGAAAAAATCTGAATATATCAATTCGTTAAGATGCGTGAATGGAAC 294
DB 2000 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
QY 295 CTAGTGGTATGAGGTTATCTCAAGACGCAAGCAACCTTTGAAA-----AAC 345
DB 2018 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 346 TTTTGGACATCAAGAAAGGATGAAATGCAAAATTCGAAATTCATTAACATCAAT 405
DB 2035 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 406 AATGACCTCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 450
DB 2055 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 451 -----AAGCAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 480
DB 2075 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 481 CCCATTACGCTCAAGTGTAAAGTAAATGCAAAATTCGAAATTCATTAACATCAAT 540
DB 2095 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 541 GCAAGTGTGGTGTGTTGTAAGGCAAGCGGCTTCCTTTGAAAGGATGATTA 600
DB 2112 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 601 GTTATGATCTCAGT-----AGGTAGGCGCTTCAAGGAT----- 636
DB 2126 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 637 -----CAGAAATGCGCTCAAGAAATTCATCAATTCATCAAGCGGTATCA 681
DB 2146 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
RESULT 9
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solnicki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
```

```
APPLICANT: Dunn, John J.
TITLE OF INVENTION: Recombinant Constructs of Borrelia
FILE REFERENCE: 2631.1001-011
CURRENT APPLICATION NUMBER: US/11/196,475
PRIOR FILING DATE: 2005-08-03
PRIOR APPLICATION NUMBER: US 08/148,191
PRIOR FILING DATE: 1993-11-01
PRIOR APPLICATION NUMBER: US 08/235,836
PRIOR FILING DATE: 1994-04-29
PRIOR APPLICATION NUMBER: US 09/666,017
PRIOR FILING DATE: 2000-09-19
PRIOR APPLICATION NUMBER: US 60/226,484
PRIOR FILING DATE: 2000-08-18
PRIOR APPLICATION NUMBER: PCT/US01/24736
NUMBER OF SEQ. ID NOS: 213
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ. ID NO 68
LENGTH: 693
TYPE: PRT
ORGANISM: Borrelia burgdorferi
US-11-196-475-68
```

Alignment Scores:

Pred. No.: 0.0819 Length: 693
Score: 117.00 Matches: 57
Percent Similarity: 41.8% Conservative: 53
Best Local Similarity: 21.7% Mismatches: 91
Query Match: 10.0% Indels: 62
Gaps: 12

US-09-360-685c-26 (1-685) x US-11-196-475-68 (1-693)

```
QY 19 GATTTACGACGATTAACGCAAGCAAAAGGACCTTGAAATTC-----GTTAAAGAT 72
DB 301 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 73 GTGATCATCATCAAAAGGTAACGATTAAGTGAATTCATCAATCAACGCGTCACTG 132
DB 321 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 133 GCTAAGCAACGCGTATTCATGAGG-----GTAGACGACGTTAGCCGAT 180
DB 341 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 169 GCGTTACCGATCTCAAAATTCGAAAGGCAATTCGCGCCACGACGTTCAAAAT 228
DB 360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 229 GAAAGTGTGATGTAAGAAATCTGAATATATCAATTCGTTAAGAT----- 279
DB 376 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 280 -----GTTGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 318
DB 396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 319 GCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 378
DB 414 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 379 GTTGAATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 438
DB 434 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 439 GTTATTAATAAGCAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 450
DB 451 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 499 GCTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 555
DB 467 -----GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
```


; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/318,700
 ; PRIOR FILING DATE: 2001-09-12
 ; NUMBER OF SEQ ID NOS: 227
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 48
 ; LENGTH: 3712
 ; TYPE: PRT
 ; ORGANISM: *Drosophila melanogaster*
 US-11-019-711-48

Alignment Scores:
 Pred. No.: 0.13 Length: 3712
 Score: 114.00 Matches: 67
 Percent Similarity: 35.0% Conservative: 46
 Best Local Similarity: 20.7% Mismatches: 100
 Query Match: 9.7% Indels: 110
 DB: 7 Gaps: 15

US-09-360-685C-26 (1-685) x US-11-019-711-48 (1-3712)

```

QY 13 AATAGGATTTTCAGCAAGTAAAGCAAGCAAAAGCGACTTGAATAATTCCTTAAGAT 72
DB 2348 AmsSerlySpheAerThrvAlserGlnLylLeuGlnAlaGlnLylSerAenileYAsp 2367
QY 73 GTG-----ATCATCAAT-----CAAAAGCTACGGATTAAGTTGAT 108
DB 2368 AlaGlyAenPheLeuileAenglyAspLeuThrvLeuAenglnileAenglnlyLeuAsp 2387
QY 109 AATCTCAATCAAGCGGATCA-----ACCTAGTCGGTATGAG----- 129
DB 2388 AenLeuAerAerAlaLeuAenGlnLeuAenSerPheAenlyAenValAerGlnLylLeu 2407
QY 130 -----GTGCTTAAGCAAGCGGATTTCTAGTAGGCTAGAG 165
DB 2408 ProValAerGlnAerGlnileYAspGlnLylAerPAlaLeuThrvAsp-----GlnAlaGln 2425
QY 166 CAAGCGTTAGCCCAATCTCAAAATTTCTCAAGCAAGCAATGGCCCAAGCAAGCTCAAAA 225
DB 2426 GlnlyValAlaAerLeuAenAlaileYAspAlaGlnAerLeuAlaAlaGlnlyThrvAspMet 2445
QY 226 AATGAAAGTCTCAAT---CTAGAAAATAATCTGAATATATCAATCCCTTAAGAAATGT 282
DB 2446 ThrAlaSerAlaGlnProAlaileYAspAlaAlaThrAlaTySerGlyileValAla 2465
QY 283 CTGAATGGA-----ACCTAGTCGGTATGAG----- 309
DB 2466 ValGlnAlaAlaGlnLylLeuSerGlnAerPAlaileSerAlaAlaGlnAerAlaThrvAsp 2485
QY 310 -----TTATCT 315
DB 2486 LyeThrvAerGlnileGlnLylAerGlnAlaileYAspAlaAerThrvGlySerThrvAerPAla 2505
QY 316 CAAGCAAGCAAGCAAGCTTTCTTAATAATTTCTGAGCAAGCAAGAAAGTTGAATGAG 375
DB 2506 GlnAerGlnAerGlnSerleuGlnileYAspAlaGlnAerPLeuGlnProAerGlnAerAla 2525
QY 376 -----AACTGGAATATTCATTAACAATAACAATAAGTCAAA 417
DB 2526 SerAlaGlnlyValGlnLylleSerAlaValAenAenAlaThrvGlnileGlnLylLeuYs 2545
QY 418 AAGCAAGCAAGCTTATGCTTAAGTTAATA----- 447
DB 2546 Asp-----IleAenlyValleuileAerGlnLylLeuProAlaGlnSerGln 2559
QY 448 -----AAGAAAAGCAAGCGCAAGCAAGCTTGA----- 477
DB 2560 ArgAspMetTrpLylAenSerAenAlaAenAlaSerAerPAlaLeuGlnLylLeuLylAaen 2579
QY 478 -----GAACCATTTAGCTCAAGTTGCTTAAG----- 507
DB 2580 ValLeuGlnLylLeuGlnLylProValSerValGlnThrvProYAspGlnLylLeuAlaAen 2599
  
```

```

QY 508 ---GTAAATGCAAAATTTGACGAGCTCAATCA-----ATAGCAAGTGTTCGGTGTGTA 561
DB 2600 GlnlyleAenAlyAerPLeuAerThrvAenlyAerPAlaAerGlnAlaAenlyGlnLylLeu 2619
QY 562 GGGCAAGCAAGCGGCTTCCTTGAAGCAAGCAAGTGAATGATGATCTCACT----- 615
DB 2620 AspAerPAlaGlnLylSerValSerlySerlyAenGlnLylLeuAlaGlnAerPLeuGlnLyl 2639
QY 616 -----AAGTGAAGGCTTCAAGCAATCAAGTGAATGCTCAAGAAATTTGCAATCTCAAT 669
DB 2640 GlnHleAerValGly---SerGlnSerAerGlnLylGlnLylleGlnLylleGlnAenlyYs 2658
QY 670 CAAGCGCTA 678
DB 2659 AlaGlnVal 2661
  
```

RESULT 12

```

US-10-793-626-2586
; Sequence 2586, Application US/10793626
; Publication No. US2005025547B1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2586
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2586
  
```

Alignment Scores:
 Pred. No.: 0.176 Length: 611
 Score: 112.50 Matches: 63
 Percent Similarity: 38.9% Conservative: 37
 Best Local Similarity: 24.5% Mismatches: 91
 Query Match: 9.6% Indels: 66
 DB: 6 Gaps: 12

US-09-360-685C-26 (1-685) x US-10-793-626-2586 (1-611)

```

QY 40 GCAAAAGCAAGCTTGAATAATTCGTTAAGATGTGATCAATCAAAAGGTAAAGGAT 99
DB 163 AlaLylAerAerPAlaileYAspThrvile---AerValleuValileGlnLylleYAsp 181
QY 100 AAGTTATATCTCAATCAAGCCGATCAAGTGGCTAA-----GCAACGGGATGATTC 153
DB 182 ---ileAerPAlaAenGlnLylLeuThrvSerGlnAerGlnAerPAlaileYAspAlaGlnLyl 200
QY 154 ACTAGGCTAGCAAGCGTTCAGCAATTCGCAAAATTTCTCAAG----- 198
DB 201 GlnAerGlnLylleGlnAlaileAerPAlaileAerPAlaileAerPAlaileAerPAlaile 220
QY 199 -----GAGCAATTCGCGCAAGCAAGCTTGAATAATTTCTCAAG----- 234
DB 221 GlnThrvAlaLylAerGThrvAerPheGlnLylleAerGlnLylleAerPAlaileAerPAlaile 240
QY 235 CTCAATCTTGAATAAATA-----TCGAAATATATCAATCCGTTAAGATGCTGTG 285
DB 241 LeuAenlyAlyLylleAerPleileThrvAerPAlaAenThrvGlnleGlnAenGlnlyPne 260
QY 286 AATGAAAGCTTGTGTTAATGCTTATCTCAAGCAAGCAAGCAAGCAAGCTTTTCAAAAC 345
DB 261 LylleGlnLylleGlnLylleYAspGlnLylleThrvAerPAlaileAerPAlaileAerPAlaile 280
  
```

QY 346 TTTTCGACATCAAGAA-----GAG 366
 Db 281 LeuThralaleuGlnlySGluPheleuGlnlyValGlnlyIleuValGlnly 300
 QY 367 TTGAAATGCAAACTTGGAAATTTCAATTAACAATTAATGACTCAAAAGCAACC 426
 Db 301 LeuasnGlnleuGlnGlnGlnGlnPheAsnAsnArgTyrGlnlyIleuAsnGlnIleu 320
 QY 427 ATT-----TATGCTAAAGTAAATTAAG 450
 Db 321 LeuGlnGlnlyGlnlyValGlnlyIleuValGlnlyValGlnlyValGlnlyValGlnly 340
 QY 451 AAAGCA-----GGGCAAGCAGCTAGCTTGAAGAACCCATTAGCTCAAGTTGCT 501
 Db 341 GlnGlnIleuAsnAsnGlnIleuValGlnlyValGlnlyValGlnlyValGlnlyValGlnly 357
 QY 502 AAAAAGTAAATGCAAAATTTGACCCGACTCAATCAATTAAGCAAGTTGGTGTGTA 561
 Db 358 -----LeuAspArgIlelyGlnIleGln-----LeuGlnIleuMet 369
 QY 562 GGGCAAGCAGCGGCTTCCTTTGAAAGCAGATGAATTAAGTTGATGCTCAAGTAAAGTA 621
 Db 370 AsnSerIleArgIly-----AlaHisThrValGlnAspValHisIleVal 384
 QY 622 GGGCTTTCAAGATCAAGATTTGGCTCAGAAATTTGACATCTCAATC 670
 Db 385 LeuLeu-GlnGlnlyIleGlnGlnIleu-----LysValAsnValSerIle 399

RESULT 13
 US-11-019-711-51
 ; Sequence 51, Application US/11019711
 ; Publication No. US2006009634A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Alsobrook II, John P
 ; APPLICANT: Tchernyev, Velizar T
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Spytak, Kimberly A
 ; APPLICANT: Baturajan, Meera
 ; APPLICANT: Grose, William M
 ; APPLICANT: Lepley, Denise M
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Vernet, Corinne A.M.
 ; APPLICANT: Li, Li
 ; APPLICANT: Gorman, Linda
 ; APPLICANT: Edinger, Shlomit R
 ; APPLICANT: Sciorio, Paul
 ; APPLICANT: Eilerman, Karen
 ; APPLICANT: Malyanar, Uriel M
 ; APPLICANT: Rothenberg, Mark
 ; APPLICANT: Stone, David J
 ; APPLICANT: Boldog, Ferenc L
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Shenoy, Suresh G
 ; APPLICANT: Anderson, David W
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Taupier ur, Raymond J
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Eisen, Andrew J
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-235
 ; CURRENT APPLICATION NUMBER: US/11/019,711
 ; PRIOR FILING DATE: 2004-12-21
 ; PRIOR APPLICATION NUMBER: US/10/037,417
 ; PRIOR FILING DATE: 2002-09-20
 ; PRIOR APPLICATION NUMBER: 60/260,018
 ; PRIOR FILING DATE: 2001-01-05
 ; PRIOR APPLICATION NUMBER: 60/260,360
 ; PRIOR FILING DATE: 2001-01-08
 ; PRIOR APPLICATION NUMBER: 60/272,411
 ; PRIOR FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: 60/272,817
 ; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/291,186
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: 60/303,231
 ; PRIOR FILING DATE: 2001-07-05
 ; PRIOR APPLICATION NUMBER: 60/305,060
 ; PRIOR FILING DATE: 2001-07-12
 ; PRIOR APPLICATION NUMBER: 60/318,405
 ; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/318,700
 ; PRIOR FILING DATE: 2001-09-12
 ; NUMBER OF SEQ ID NOS: 227
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 51
 ; LENGTH: 3712
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-11-019-711-51

Alignment Scores:
 Pred. No.: 0.182 Length: 3712
 Score: 112.00 Matches: 68
 Percent Similarity: 33.7% Conservative: 41
 Best Local Similarity: 21.1% Mismatches: 104
 Query Match: 9.6% Indels: 110
 DB: 7 Gaps: 14

US-09-360-685C-26 (1-685) x US-11-019-711-51 (1-3712)

QY 13 AATAAGATTTCAGACAGTAAGCAAGCAAAAGCAGCTTGAATAATTCCTTAAGAT 72
 Db 2348 AsnSerIlePheAsnThrValSerGlnlySGlnlyValGlnlyValGlnlyValGlnly 2367
 QY 73 GTG-----ATCATCAAT-----CAAAAGTAAAGTAAAGTTGAT 108
 Db 2368 AlaGlnAsnPheLeuIleAsnGlnlyAspLeuThrLeuAsnGlnIleAsnGlnlyLeuAsp 2387
 QY 109 AATTCATCAAGCGGTATCA----- 129
 Db 2388 AsnLeuArgAspAlaLeuAsnGlnlyLeuAsnSerPheAsnLysAsnValAspGlnlyLeu 2407
 QY 130 -----GTGGCTAAAGCAAGCGGTATTCAGTAGGCTAGAG 165
 Db 2408 ProValArgGlnAspGlnIleLysGlnlyValAspAlaThrIleAsp-----GlnAlaGln 2425
 QY 166 CAAGCGTTAGCGCATCTCAAAATTTCTCAAGAGCAATTGCCCAACAAGCTCAAAA 225
 Db 2426 GlnlyValAlaGlnLeuAlaIleLysAlaGlnAspLeuAlaAlaGlnIleThrAspMet 2445
 QY 226 AATGAAGTCTCAAT--GCTAGAAAATAATCTGAATATATCAATCCGTTAAGATGCT 282
 Db 2446 ThrAlaSerAlaGlnProAlaIleLysAlaAlaThrAlaIleAspGlnlyIleValGlnly 2465
 QY 283 GTGAATGGA-----ACCTAGTCGTTAATGG----- 309
 Db 2466 ValGlnAlaAlaGlnlyLeuSerGlnAspAlaIleSerAlaValAlaGlnlyValThrAsp 2485
 QY 310 -----TTACT 315
 Db 2486 LysThrAspGlnlyIleGlnGlnArgAlaHisLeuAlaAspThrGlnlySerThrAspLeu 2505
 QY 316 CAAGCAGAGCCACACTTTCTTAAAACTTTGCGACCAAGAAAGCTTAATGCA 375
 Db 2506 GlnArgAlaArgGlnSerLeuGlnlyValGlnAspLeuGlnlyProArgLeuAsnAla 2525
 QY 376 -----AACTTGAATTTCAATTAACAATTAATGACTCAAA 417
 Db 2526 SerAlaGlnlyValGlnlyValSerAlaValAsnAsnAlaThrGlnlyIleGlnly 2545
 QY 418 AACGAACCATTTATGTAAAGTTAATAA----- 447
 Db 2546 Asp-----IleAsnlyLeuIleAspGlnlyProAlaGlnlySerGln 2559
 QY 448 -----AGAAAGCAGGGCAGACAGCTAGCTTGA----- 477

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 6, 2006, 20:04:12 ; Search time 28.7 Seconds
(without alignments)

1994.516 Million cell updates/sec

Title: US-09-360-685c-26

Perfect score: 1171

Sequence: 1 aaatgcgaataataga.....caatcagcgatcatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_n2p.model -DEV=rlp
-O=/abs/ABSSWEB_spool/US09360685/runat_06032006_121327_15626/app_query.fasta_1
-DB=Published Applications AA Main -QFMT=fasta -SUPFIX=rapbm -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blomsu62
-TRANS=huma40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -HOST=abs802p
-USER=US09360685@CGN_1_1_307@runat_06032006_121327_15626 -NCPU=6 -ICPU=3
-NO_WMAP -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA Main:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	3	US-09-921-157-5
2	1127	96.2	1338	3	US-09-402-100-4
3	896.5	76.6	1167	4	US-10-335-977-8633
4	896.5	76.6	1178	4	US-10-335-977-8634
5	896.5	76.6	1183	4	US-10-335-977-8635
6	140	12.0	6713	4	US-10-282-122A-43811
7	131.5	11.2	1404	5	US-10-732-923-3304
8	131	11.2	3533	4	US-10-282-122A-70177
9	129	11.0	1992	5	US-10-470-048B-81
10	129	11.0	10498	5	US-10-470-048B-440
11	127	10.8	837	3	US-09-815-242-5883

12	127	10.8	875	3	US-09-815-242-13080	Sequence 13080, A
13	127	10.8	1361	4	US-10-369-493-3209	Sequence 3209, Ap
14	127	10.8	2434	3	US-09-815-242-5835	Sequence 5835, Ap
15	127	10.8	6281	3	US-09-815-242-12996	Sequence 12996, A
16	126	10.8	2368	3	US-09-815-242-5635	Sequence 5635, Ap
17	126	10.8	2368	3	US-09-815-242-12389	Sequence 12389, A
18	125.5	10.7	2437	3	US-09-815-242-5834	Sequence 5834, Ap
19	125	10.7	2478	3	US-09-815-242-5816	Sequence 5816, Ap
20	125	10.7	2478	3	US-09-815-242-12967	Sequence 12967, A
21	125	10.7	2478	5	US-10-470-048B-220	Sequence 220, App
22	125	10.7	2481	4	US-10-282-122A-43762	Sequence 43762, A
23	124	10.6	5795	3	US-09-815-242-12610	Sequence 12610, A
24	123	10.5	10203	4	US-10-661-809-23	Sequence 23, Appl
25	123	10.5	10203	4	US-10-724-972A-4098	Sequence 4098, Ap
26	122.5	10.5	998	4	US-10-282-122A-70540	Sequence 70540, A
27	121.5	10.4	6641	4	US-10-282-122A-70560	Sequence 70560, A
28	121	10.3	993	4	US-10-282-122A-43875	Sequence 43875, A
29	120	10.2	724	4	US-10-282-122A-71401	Sequence 71401, A
30	120	10.2	1090	4	US-10-282-122A-56217	Sequence 56217, A
31	120	10.2	3158	3	US-09-815-242-12611	Sequence 12611, A
32	118	10.1	1441	5	US-10-732-923-3352	Sequence 3352, Ap
33	118	10.1	1693	4	US-10-603-725-4	Sequence 4, Appl1
34	118	10.1	1693	4	US-10-603-725-8	Sequence 8, Appl1
35	118	10.1	1713	4	US-10-171-311-113	Sequence 113, App
36	118	10.1	1713	4	US-10-372-683-10	Sequence 10, Appl
37	118	10.1	1713	4	US-10-603-725-6	Sequence 6, Appl1
38	118	10.1	1724	4	US-10-603-725-2	Sequence 2, Appl1
39	117.5	10.0	1047	4	US-10-282-122A-44353	Sequence 44353, A
40	117.5	10.0	2025	3	US-09-815-242-5703	Sequence 5703, Ap
41	117	10.0	693	4	US-10-369-100-68	Sequence 68, Appl
42	116	9.9	2067	6	US-11-097-143-40167	Sequence 40167, A
43	116	9.9	2539	5	US-10-831-070-6	Sequence 6, Appl1
44	115	9.8	873	3	US-09-952-267-13	Sequence 13, Appl
45	115	9.8	873	5	US-10-872-768-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-921-157-5
Sequence 5, Application US/09921157
Publication No. US20040048353A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonio
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappelli, Rino
TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For
FILE REFERENCE: CHIR0315
CURRENT APPLICATION NUMBER: US/09/921,157
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/466,662
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/360,934
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: PCT/EP93/00472
PRIOR FILING DATE: 1993-03-02
PRIOR APPLICATION NUMBER: PCT/EP93/00158
PRIOR FILING DATE: 1993-01-25
PRIOR APPLICATION NUMBER: FI 92 A 000052
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori

US-09-921-157-5

Alignment Scores:

Pred. No.: 1,3e-96
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 96.2%
DB: 3

Length: 1147
Matches: 228
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-360-685C-26 (1-685) x US-09-921-157-5 (1-1147)

```
QY 1 AAAAAAAAAATTAAGATTTCAGCAAGTAAAGCAAGCAAAAGCACTTGAAT 60
DB LysanGlyLysanLysanLysanPheSerLysValThrGlnLalysSerPheLysGln 769
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGCAAGTAAAGTAAATCTCA 120
DB SerValLysanPValIleIleAnGlnLysValThrAspLysValAspAnLysanGln 789
QY 121 GCGGTATCAAGTGTCAAGCAAGCGGTGATTCAGTGGGTAGCAAGCGTTAGCC 180
DB AlValSerValAlalysAlatnGlyAspPheSerLysValGlnGlnAlalValAsp 809
QY 181 CTCAAAATTTCTCAAGAGCAATGGCCCAACAACTCAAAAATGAAAGTCTCA 240
DB LeuLysanPheSerLysGlnGlnLeuAlaGlnGlnLysanGlnSerLysan 829
QY 241 GCTAGAAAAAATCTGAATAATATATCAATCGTTAAGATGCTGAATGAAACCT 300
DB AlaArgLysLysSerGlnLysValLysanGlnLysanGlnLysanGlnLysan 849
QY 301 GGTATAGGTTATCTCAAGAGCAAGCAAGCTTTCTAAAACCTTTGGACATCA 360
DB 850 GlyAnGlyLysSerGlnLysanLysanLysanLysanLysanLysanLysan 869
QY 361 AAGAGTTGAATGCAAACTTGAATTTCAATTAACATTAATTAATGACTCAAA 420
DB LysGlnLysanLalysLysanLysanPheSerLysanLysanLysanLysan 889
QY 421 GAACCATTTATGCTAAAGTAAATTAAGCAAGCGCAAGCGTACCTTGAAGAA 480
DB GlnProLysLysanLysanLysanLysanLysanLysanLysanLysanLysan 909
QY 481 CCATTTAGCTCAAGTGTCTAAAAGTAAATGCAAAATTTGACGATCAATCAAT 540
DB 910 ProLysLysanLysanLysanLysanLysanLysanLysanLysanLysan 929
QY 541 GCAAGTGTGGTGTCTTGAAGCAAGCGGCTTCCCTTTGAAAGGATGATTA 600
DB 930 AlaSerGlnLysanLysanLysanLysanLysanLysanLysanLysanLysan 949
QY 601 GTTATGATCTCAGTAAAGTGGCTTTCAAGGAATCAAGAAATGGCTCAAGAA 660
DB 950 ValAspPheSerLysanLysanLysanLysanLysanLysanLysanLysanLysan 969
QY 661 AATTCATCAAGCGGTATCAGAA 684
DB 970 AsnLysanGlnLalValSerGln 977
```

RESULT 2

US-09-402-100-4

Sequence 4, Application US/09402100

Patent No. US20010019834A1

GENERAL INFORMATION:

Applicant: Daewoong Pharmaceutical Co., LTD

Applicant: Kim, Byung-O

Applicant: Shin, Sung-Seup

Applicant: Yu, Young-Hyo

Applicant: Park, Myung-Hwan

Applicant: Choi, Deok-Joon

Applicant: Jung, Hyung-Jin

TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Helicob

FILE REFERENCE: 0136/003140
CURRENT APPLICATION NUMBER: US/09/402,100
CURRENT FILING DATE: 1999-09-27
EARLIER APPLICATION NUMBER: KR 97-11950
EARLIER FILING DATE: 1997-03-31
EARLIER APPLICATION NUMBER: KR 97-11951
NUMBER OF SEQ ID NOS: 48
SOFTWARE: Patentin version 3.0
SEQ ID NO: 4
LENGTH: 1338
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Caga/CTA2B Chimeric protein
US-09-402-100-4

Alignment Scores:

Pred. No.: 1.32e-96
Score: 1127.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 96.2%
DB: 3

Length: 1338
Matches: 228
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-360-685C-26 (1-685) x US-09-402-100-4 (1-1338)

```
QY 1 AAAAAAAAAATTAAGATTTCAGCAAGTAAAGCAAGCAAAAGCACTTGAAT 60
DB LysanGlyLysanLysanLysanPheSerLysValThrGlnLalysSerPheLysGln 769
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGCAAGTAAAGTAAATCTCA 120
DB SerValLysanPValIleIleAnGlnLysValThrAspLysValAspAnLysanGln 789
QY 121 GCGGTATCAAGTGTCAAGCAAGCGGTGATTCAGTGGGTAGCAAGCGTTAGCC 180
DB AlValSerValAlalysAlatnGlyAspPheSerLysValGlnGlnAlalValAsp 809
QY 181 CTCAAAATTTCTCAAGAGCAATGGCCCAACAACTCAAAAATGAAAGTCTCA 240
DB LeuLysanPheSerLysGlnGlnLeuAlaGlnGlnLysanGlnSerLysan 829
QY 241 GCTAGAAAAAATCTGAATAATATATCAATCGTTAAGATGCTGAATGAAACCT 300
DB AlaArgLysLysSerGlnLysValLysanGlnLysanGlnLysanGlnLysan 849
QY 301 GGTATAGGTTATCTCAAGAGCAAGCAAGCTTTCTAAAACCTTTGGACATCA 360
DB 830 GlyAnGlyLysSerGlnLysanLysanLysanLysanLysanLysanLysan 869
QY 361 AAGAGTTGAATGCAAACTTGAATTTCAATTAACATTAATTAATGACTCAAA 420
DB LysGlnLysanLalysLysanLysanPheSerLysanLysanLysanLysan 889
QY 421 GAACCATTTATGCTAAAGTAAATTAAGCAAGCGGCAAGCGTACCTTGAAGAA 480
DB GlnProLysLysanLysanLysanLysanLysanLysanLysanLysanLysan 909
QY 481 CCATTTAGCTCAAGTGTCTAAAAGTAAATGCAAAATTTGACGATCAATCAAT 540
DB 910 ProLysLysanLysanLysanLysanLysanLysanLysanLysanLysan 929
QY 541 GCAAGTGTGGTGTCTTGAAGCAAGCGGCTTCCCTTTGAAAGGATGATTA 600
DB 930 AlaSerGlnLysanLysanLysanLysanLysanLysanLysanLysanLysan 949
QY 601 GTTATGATCTCAGTAAAGTGGCTTTCAAGGAATCAAGAAATGGCTCAAGAA 660
DB 950 ValAspPheSerLysanLysanLysanLysanLysanLysanLysanLysanLysan 969
QY 661 AATTCATCAAGCGGTATCAGAA 684
DB 970 AsnLysanGlnLalValSerGln 977
```

Db 970 AsnLeuAsnGlnAlaValSerGlu 977

RESULT 3
US-10-335-977-8633
Sequence 8633, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:
NAME: Mandagouras, Amy E.
REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8633:
SEQUENCE CHARACTERISTICS:
LENGTH: 1167 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori

FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1167
SEQUENCE DESCRIPTION: SEQ ID NO: 8633:
US-10-335-977-8633

Alignment Scores:
Pred. No.: 4.65e-75 Length: 1167
Score: 896.50 Matches: 189
Percent Similarity: 74.8% Conservative: 7
Best Local Similarity: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
DB: Gaps: 2

US-09-360-685C-26 (1-685) x US-10-335-977-8633 (1-1167)

QY 1 AAAAATGCGAAATAATGAGATTTCGACGAGTAAGCAACGAAAGCGACTTGAAAT 60
Db 754 LysAsnGlyLysAsnLysAspPheSerValThrGlnAlaLysSerAspLeuGln 773
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGTAAGCAATGATGATATCTCATGAA 120
Db 774 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspAsnLeuAsnGln 793
QY 121 GCGGATACGTGGCTTAAGCAACGGGTGATTCAGTAGGGTAGCAACGCTTAGCCGAT 180
Db 794 AlaValSerValAlaLysAlaThrIleLysAspPheSerGlyValGlnGlnAlaLeuAlaAsp 813

QY 181 CTCAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAAATGAATCTCAAT 240
Db 814 LeuLysAsnPheSerLysGlnGlnLeuAlaGlnGlnAlaGlnLysAsnGlnAspPheAsn 833
QY 241 GCTAGAAAAAATATGAAATATATCAATCCGTTAGAAATGGTGCAATGAAACCTAGTC 300
Db 834 ThrGlyLysAsnSerAlaLeuIleGlnSerValIleAsnGlyValAsnGlyThrLeuVal 853
QY 301 GGTAAATGGTTATCTCAAGACAGAACCAACTTTCTTAAAACTTTCCAGATCAAG 360
Db 854 GlyAsnGlyLeuSerLysAlaGlnAlaThrThrIleuSerLysAsnPheSerAspIleLys 873
QY 361 AAAGGTGAATGCAAAATCTTGAAATTTCAATTAACATTAACATTAATGACTCAAAAAC 420
Db 874 LysGlnLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuGlnAsn 893
QY 421 GAACCATTTATGCTAAAGTTAATTAAGAAACAGGCGCAAGCAGTACCTTGAAAGA 480
Db 894 Ser-----ThrGln 896
QY 481 CCCATTATGCTCAAGTTGCTTAAAGGTAATGCAAAATTTGACCGACTCAATCAATA 540
Db 897 ProIleTyThrGlnValAlaLysLysValLysAlaLysIleAspArgLeuAspGlnIle 916
QY 541 GCAAGTGTGGGTGGTGTGAGGCGCAAGACGGGC----- 576
Db 917 AlaSerGlyLeuGlyAspValGlyGlnAlaLaserPheLeuLysArgHisAspLys 936
QY 576 ----- 576
Db 937 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProIleTyAlaThrIle 956
QY 577 -----TTCCCTTGAAGGATGATTAAGTTGATGATTCAGTAAG 618
Db 957 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSerLys 976
QY 619 GTAGGGCTTTCAAGGAATCAAGAAATGGCTCAGAAATTTGACATTCATCAAGCGGTA 678
Db 977 ValGlyLeuSerArgLysGlnLysLeuThrGlnLysIleAspAsnLeuAsnGlnAlaVal 996
QY 679 TCAGAA 684
Db 997 SerGlu 998

RESULT 4
US-10-335-977-8634
Sequence 8634, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

US-09-360-685c-26 (1-685) x US-10-335-977-8635 (1-1183)

```

QY 1 AAAAAAGGCAAAATAGATTTTACGACGAGTACGCAAGCAAAAGCCGCTTGAAT 60
Db 770 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspLeuGln 789
QY 61 TCCGTTAAAGATGTGATTCATCAATCAAAAGGTAAAGGATTAAGTTATATTCATCA 120
Db 790 SerLleLysAspValLleIleAsnGlnLysLleThrAspLysValAspAsnLeuAsn 809
QY 121 GCGGATTCAGTGGCTTAAGCAAGCGGTGATTTGATGAGGTAGAGCAAGCGTAC 180
Db 810 AlaValSerValAlaLysAlaThrGlyAspPheSerGlyValGluGlnAlaLeuAla 829
QY 181 CTCGAAATTTCTCAAGAGCAATTTGGCCCAAGCTTCAGCAAAATGAAAGTCTCAT 240
Db 830 LeuLysAsnPheSerLysGluGlnLeuAlaGlnGlnAlaGlnLysAsnGluAspPhe 849
QY 241 GCTAGAAAAATCTGAATATATATATATATATATATATATATATATATATATAT 300
Db 850 ThrGlyLysAsnSerAlaLeuLysValLysAsnGlyValAsnGlyThrLeuVal 869
QY 301 GGTATGCGTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
Db 870 GlyAsnGlyLeuSerLysAlaGluAlaThrThrLeuSerLysAsnPheSerAspLle 889
QY 361 AAGAGTTGATGCAAACTTGAATTTTCATTAACATTAACATTAACATTAACATTA 420
Db 890 LysGluLeuAsnAlaLysLeuGlyAsnPheAsnAsnAsnAsnAsnLysLeuGluAsn 909
QY 421 GAACCATTTATGCTAAAGTTATATAAAGAAAGCAAGCAAGCAAGCAAGCAAGCA 480
Db 910 Ser-----ThrGlu 912
QY 481 CCCATTTAGCTCAAGCTGCTAATAAAGGTAATGCAAAATTTGACGCACTCAATCA 540
Db 913 ProLleLysThrGlnValAlaLysLysValLysAlaLysLleAspArgLeuAsnGln 932
QY 541 GCAAGTGGTTGGGCTTGTGAGGCAAGCAAGCGGCGC----- 576
Db 933 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuLysArgHisAsp 952
QY 576 ----- 576
Db 953 ValAspAspLeuSerLysValGlyLeuSerAlaAsnHisGlnProLleLysAlaThr 972
QY 577 -----TTCCCTTTGAAAGGCAATGATTAAGTTGATGATGATGATGATGATG 618
Db 973 AspAspLeuGlyGlyProPheProLeuLysArgHisAspLysValAspAspLeuSer 992
QY 619 GTAGGCTTTCAAGCAATCAAGATTTGGCTCAAGAAATTTGACAAATCTCAATCAAG 678
Db 993 ValGlyLeuSerArgLysGlnLysLeuThrGlnLysLleAspAsnLeuAsnGlnAla 1012
QY 679 TCAGAA 684
Db 1013 SerGln 1014

```

RESULT 6
US-10-282-122A-43811
; Sequence 43811, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Tiawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 43811
LENGTH: 6713
TYPE: PRT
ORGANISM: *Staphylococcus aureus*
US-10-282-122A-43811

Alignment Scores:
Pred. No.: 0.00033 Length: 6713
Score: 140.00 Matches: 66
Percent Similarity: 38.8% Conservative: 50
Best Local Similarity: 22.1% Mismatches: 97
Query Match: 12.0% Indels: 86
DB: Gaps: 11

US-09-360-685c-26 (1-685) x US-10-282-122A-43811 (1-6713)

```

QY 13 AATAAGATTTACAGCAAGTAAAGCAAGCAAAAGCCAGC----- 51
Db 5825 AsnLysAspValAspLysGlnValGlnAlaLeuLleAspGluLleAspArgAsnProAsn 5844
QY 52 CTGAAATTCGCTTAAGATGTGATC-----ATCAATCAAAAGTAAAGCAATTA 102
Db 5845 LeuThrAspLysGluLysGlnAlaLeuLysAspArgLleAsnGlnLleLeuGlnGln 5864
QY 103 GTGATATCTCATCAAGCGGTATCAAGTGGCTTAAGCAAGCGGTGATTTCAAGTAG 162
Db 5865 HisAsnAspLleAsnAsnAlaLeuThrLysGluGlnLleGluGlnAlaLysAlaGln 5884
QY 163 GAGCAAGCTTAGCGCATCTCAAAATTTCTCAAG----- 198
Db 5885 AlaGlnAlaLeuGlnAspLleLysAspLeuValLysAlaLysGluAspAlaLysGlnAsp 5904
QY 199 -----GAGCAATTTGGCCCAAGCAAGCTCAAAATTTGAGTCAATGCT 243
Db 5905 ValAspLysGlnValGlnAlaLeuLleAspGluLleAspGlnAsnProAsnLeuThrAsp 5924
QY 244 AGAAAAA-----TCTGAATATATATATATATATATATATATATATATATATAT 282
Db 5925 LysGluLysGlnAlaLeuLysAspArgLleAsnGlnLleLeuGlnGlnGlnLysAla 5944
QY 283 GTGAATGAACCCAGTCGGAATAGGTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCT 342
Db 5945 IleAsnAsnAlaAsnThrLysGluGluLleGluGlnAlaLysAla-----GlnLeuAlaGln 5963

```

Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbm

Page 6

```
QY 343 AACTTTGGACATCAG----- 360
Db 5964 AAlaleuylsegilullelyshapleuvalylsalylsgluamalaelyglnaspaalasp 5983
QY 361 AAAGAGTTGAATGCAAAACTGGAAATTCATTAACATTAACAT----- 405
Db 5984 Lyeginvalglmlaleuileasglulleaspglnamprokashenuthrasplyeglu 6003
QY 406 -----AATGACTCAAA 417
Db 6004 Lyeginlaleuylshapargyllaenglnilleuglnlnclylshasnaasplleas 6023
QY 418 AACGACCCATTATGCTTAAGTTAAATAAAGAACGAGGACGAGCTAGCTTGA 477
Db 6024 Asnalamecthrlyegluluglulleglulnalaalsaglmlaleuialaleugln 6043
QY 478 GAACCCATTAGCTCAAGTTGCTAAAGGTAATGCAAAATGACCGACTCAAT--- 534
Db 6044 Asp-----lilelysaaspleuvallylsalalygluaaspaalalyasnaala 6058
QY 535 -----CAATAGCAAGTGGCTTGGCTGGCTGTTAGGG 564
Db 6059 lilelyalaleuialasnaalalyshaparglnilleasnsasnaasprokashenuthrpro 6078
QY 565 CAAGCAGCGGCTTCCCTTGAAGAGCATGATTAAGTTGATGATCTCAGTAAAGG 624
Db 6079 Gluglnlyalalyalalyaleuyls-----glulleaspglnalaglnulyshargla 6095
QY 625 CTTTCAGGAATCAAGATTGGCTCAAGAAATTTGACATCTTCATCAACGCGGTATCA 681
Db 6096 Leu---GlnasnaValGlnasnaalaglnthrilleaspglnleuamagglyleuam 6113

RESULT 7
US-10-732-923-3304
/ Sequence 3304, Application US/10732923
/ Publication No. US20050108791A1
/ GENERAL INFORMATION:
/ APPLICANT: Edgerton, Michael D
/ TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
/ FILE REFERENCE: 38-15(52796)C
/ CURRENT APPLICATION NUMBER: US/10/732,923
/ PRIOR FILING DATE: 2003-12-10
/ PRIOR APPLICATION NUMBER: 10/310,154
/ PRIOR FILING DATE: 2002-12-04
/ NUMBER OF SEQ ID NOS: 24149
/ SEQ ID NO 3304
/ LENGTH: 1404
/ TYPE: PRF
/ ORGANISM: Mycoplasma hominis
US-10-732-923-3304

Alignment Scores:
Pred. No.: 0.00164
Score: 131.50
Percent Similarity: 44.9%
Best Local Similarity: 20.6%
Query Match: 11.2%
DB: 5
Gaps: 9
```

```
US-09-360-685C-26 (1-685) x US-10-732-923-3304 (1-1404)
QY 1 AAAATGCAAAATTAAGGATTTCAGCAAGTAACGCAAGCAAAAGCACTTGAAT 60
Db 112 Lyasnaulyleylshlleglulnleaspglnlylslevalglmlameglnluphlyls 131
QY 61 TCCGTTAAAGATG-----ATCATCATCAATAAAGGTAACGATTAAGTTGATATTC 114
Db 132 Serclnlyalaleuileugllyaspleulleasnsasnaasprokashenuthr 151
QY 115 AATCAAGCGTATCACTGCTTAAGCAAGCG---CGTATTTGATAGGAGGAGCAAGC 171
Db 152 Asnalalyeglinslerleuglnasnaenthryalashasnllysserlileglnl 171
```

US-09-360-685C-26 (1-685) x US-10-732-923-3304 (1-1404)

```
QY 172 TTACCGCATCTCAAAATTTCTCAAGAGCAATGGCCCAACAGCTCAAAATATGA 231
Db 172 lileglnlalelseserlyllleasnglnlalyelyleglulleuglnserglnilleasna 191
QY 232 AGTCTCAATGCTAAGAAAAATCTGAAATATATCAATCCGTTAAGAAATGCTGAATGA 291
Db 132 Alahargasnglnlyls-----Gluvalphesglululyelyleglnlneuenly 209
QY 232 ACCCTAGTGGTAATGGCTTA-----TTCAGACGAGCAAGCAACTCTTTTAA 342
Db 210 LeuitleySerasenglnulleasnaasnerlyelalaleaspglnuthrilleu 249
QY 343 AACTTTGGACATC-----AAGAAAGAGTG 369
Db 230 AsnthrasnaValValGlyaspserrllelythrillegluthrlyserthrlyse 249
QY 370 AATGCAAACTTGAATTTCAATTAATCAATTAATGCACTCAAAAGCAACCATTTAC 429
Db 250 Glulysalalilegluserrleuthrasnlylleasnglu---phelyslyeglulnglu 268
QY 430 TATGCTAAAGTTAATAAAGAAAGCAAGGCAAGGCTGACCTGAAACCATTTAC 489
Db 269 LyalaleasnaVallylsalavalpheserlylyserlysglnleuylaspleuileasp 288
QY 490 GCTCAAGTTGCTAAAGTAATGCAAAATTTGACGACTCAATCAATTAATGCAAGTGT 549
Db 289 Sergluaspgllyelylevalaspsersasnglu---serlnvalleuthr--- 305
QY 550 TTGGGTGTGTAGGCAAGCAAGCGGCTTCCCTTGAAGAGCATGATTAAGTTGATGAT 609
Db 306 -----lyethrlylleaspgln 311
QY 610 CTCAGTAAGTGAAGG-----CTTCAAGAAATCAAGAAATTTGCAAAATTTGCAAT 663
Db 312 Asnserserillegluaspglileglnasnaenthryshasplleuilyalalegluser 331
QY 664 CTCATCAACCGTATCAAG 684
Db 332 LeuthrasnaVallylleasnaasp 338
```

RESULT 8

US-10-282-122A-70177

/ Sequence 70177, Application US/10282122A

/ Publication No. US20040029129A1

/ GENERAL INFORMATION:

/ APPLICANT: Wang, Liangsu

/ APPLICANT: Zamudio, Carlos

/ APPLICANT: Malone, Cheryl

/ APPLICANT: Haselbeck, Robert

/ APPLICANT: Ohlsen, Karl

/ APPLICANT: Zykkind, Judith

/ APPLICANT: Wall, Daniel

/ APPLICANT: Trawick, John

/ APPLICANT: Carr, Grant

/ APPLICANT: Yamamoto, Robert

/ APPLICANT: Forsyth, R.

/ APPLICANT: Xu, H.

/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

/ FILE REFERENCE: ELITRA.034A

/ CURRENT APPLICATION NUMBER: US/10/282,122A

/ PRIOR FILING DATE: 2003-02-20

/ PRIOR APPLICATION NUMBER: 60/191,078

/ PRIOR FILING DATE: 2000-03-21

/ PRIOR APPLICATION NUMBER: 60/206,848

/ PRIOR FILING DATE: 2000-05-23

/ PRIOR FILING DATE: 2000-07-727

/ PRIOR FILING DATE: 2000-05-26

/ PRIOR APPLICATION NUMBER: 60/230,335

/ PRIOR FILING DATE: 2000-09-06

/ PRIOR APPLICATION NUMBER: 60/230,347

/ PRIOR FILING DATE: 2000-09-09

/ PRIOR APPLICATION NUMBER: 60/242,578

/ PRIOR FILING DATE: 2000-10-23

Db 1286 LeuylsAspArgIleAsnGlnIleLeuGlnGlnGlyHleAsnAspIleAsnAsnAlaLeu 1305
Qy 298 GTGGTAATGGCTTATCTCAAGCAAGCCACACTCTTTCTAAAACCTTTTCGAGATC 357
Db 1306 ThrlyseGluIuileGluGlnAlaLysAla---GlnLeuAlaGlnAlaLeuGlnAspIle 1324
Qy 358 AAGAAAGATTGATGCAAACTTGAAATTCATTAACATTAACATTAATGACATCAAA 417
Db 1325 Lys----- 1325
Qy 418 AACGAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGCAAGCAAGCTTGA 477
Db 1326 ---AspleuVallyAlaLysGluAspAlaLysAsnAlaIleAlaLeuAla----- 1342
Qy 478 GAACCATTTACGCTCAAGTTGCTAAAGTAAATGCAAAATTCAGCGACTCAATCA 537
Db 1343 ----- 1343
Qy 538 ATAGCAAGTGTGGTGTGTGTAGGCAAGCGGCGCTCCCTTGAAGCAATGAT 597
Db 1351 ---SerAsnProAspleuThrProGlnGlnlyAlaLysAlaLeuLys----- 1355
Qy 598 AAGTTGATGATCTCAAGTGAAGGCTTTCAGGAATCAGAAATTCGGCTGAAATT 657
Db 1366 GluIleAspGluAlaGluLysArgAlaLeu---GlnAsnValGluAsnAlaGlnThrIle 1384
Qy 658 GACATCTCAATCAAGCGGTATCA 681
Db 1385 AspGlnLeuAsnArgGlyLeuAsn 1392
RESULT 10
US-10-470-048B-440
/ Sequence 440, Application US/10470048B
/ Publication No. US2005003744A1
/ GENERAL INFORMATION:
/ APPLICANT: MEINKE ET AL.
/ TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
/ TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
/ FILE REFERENCE: SONN-034US
/ CURRENT APPLICATION NUMBER: US/10/470, 048B
/ CURRENT FILING DATE: 2003-07-25
/ NUMBER OF SEQ ID NOS: 603
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 440
/ LENGTH: 10498
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
/ FEATURE:
/ NAME/KEY: MOD_RES
/ LOCATION: (9728)..(9731)
/ OTHER INFORMATION: X = anything
US-10-470-048B-440
Alignment Scores:
Pred. No.: 0.00377 Length: 10498
Score: 129.00 Matches: 65
Percent Similarity: 39.2% Conservative: 40
Best Local Similarity: 24.3% Mismatches: 71
Query Match: 11.0% Indels: 92
DB: Gaps: 13
US-09-360-685C-26 (1-685) x US-10-470-048B-440 (1-10498)
Qy 16 AAGATTTCAGCAAGGTAAAGCAA---GCAGAAAGCAAGCTTGAATAATCCGTTAAAGAT 72
Db 9677 LysAspleuVallyAlaLysGluAspAlaLysGlnAspValAspleuValGlnAla 9696
Qy 73 GTGATC-----ATCAATCAAAAGCTAAGCAATTA-----GTT 105
Db 9697 LeuIleAspGluIleAspGlnAsnProAsnLeuThrAspleuGluLysGlnAlaLeuLys 9716
Qy 106 GATATCTCAATCA----- 120

Db 9717 AspArgIleAsnGlnIleLeuGlnGlnGlyHleAsnIleAsnAlaLeuLys 9736
Qy 121 ---CGGTATCAAGCTTAAAGCAAGCGGTATTCAGTAGGATAGCAAGGTTAGCC 177
Db 9737 GluAlaIleGlnGlnAlaLys-----GluArgLeuAlaGlnAlaLeuGln 9751
Qy 178 GATCTCAAAATTTCTCAAG----- 198
Db 9752 AspleuValAspleuVallyAlaLysGluAspAlaLysAsnAspIleAspleuVal 9771
Qy 199 GAGCATTTGGCCCAAGCTCAAAAATGAAAGTCTCAATGCTAGAAAAA----- 252
Db 9772 GlnAlaLeuIleAspGluIleAspGlnAsnProAsnLeuThrAspleuGluLysGlnAla 9791
Qy 253 -----TCTGAATATATCAATCCGTTAAGAAATGCTGATGAAACCTTA 297
Db 9792 LeuLysAspArgIleAsnGlnIleLeuGlnGlnGlyHleAsnAspIleAsnAlaLeu 9811
Qy 298 GTGGTAATGGCTTATCTCAAGCAAGCCACACTCTTTCTAAAACCTTTTCGAGATC 357
Db 9812 ThrlyseGluIuileGluGlnAlaLysAla---GlnLeuAlaGlnAlaLeuGlnAspIle 9830
Qy 358 AAGAAAGATTGATGCAAACTTGAAATTCATTAACATTAACATTAATGACATCAAA 417
Db 9831 Lys----- 9831
Qy 418 AACGAACCATTTATGCTAAAGTTAATAAAGAAAGCAAGCAAGCAAGCTTGA 477
Db 9832 ---AspleuVallyAlaLysGluAspAlaLysAsnAlaIleAlaLeuAla----- 9848
Qy 478 GAACCATTTACGCTCAAGTTGCTAAAGTAAATGCAAAATTCAGCGACTCAATCA 537
Db 9849 ----- 9856
Qy 538 ATAGCAAGTGTGGTGTGTGTAGGCAAGCGGCGCTCCCTTGAAGCAATGAT 597
Db 9857 ---SerAsnProAspleuThrProGlnGlnlyAlaLysAlaLeuLys----- 9871
Qy 598 AAGTTGATGATCTCAAGTGAAGGCTTTCAGGAATCAGAAATTCGGCTGAAATT 657
Db 9872 GluIleAspGluAlaGluLysArgAlaLeu---GlnAsnValGluAsnAlaGlnThrIle 9890
Qy 658 GACATCTCAATCAAGCGGTATCA 681
Db 9891 AspGlnLeuAsnArgGlyLeuAsn 9898
RESULT 11
US-09-315-242-5883
/ Sequence 5883, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Resestial Genes in
/ FILE REFERENCE: ELITRA, 011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625


```

Db      |||      :|||      :|||      :|||      :|||
587 LeuAnGlyAenGluAenLeuGluAalAaLySgInGlnIaSerGlnSerLeuGlySer 606
Qy      226 AATGAAAGCTCAATGCTTAAGAAAAAATCAATATATCAATCCGTTAGAAATGCTGTG 285
Db      :|||      :|||      :|||      :|||      :|||
607 LeuAenLeuAenAenAaLySgInLySgInLySgInLySgInLySgInLySgIn 622
Qy      286 AATGCA-----GlnThrValThrAenGlnIle 291
Db      |||      :|||      :|||      :|||      :|||
623 AenGlyAalAaSerThrValAaSerGluAaenGlnIleLySgInAaenAaGlnAaenLeuAen 642
Qy      292 ACCCTTAGTCGGTAAT-----GGGTTATCTCAAGCAAGAACCAACTCTTCTMAA 342
Db      :|||      :|||      :|||      :|||      :|||
643 ThrAlaMetGlyAenLeuLySgInAlaIleAaAaSerLyAaSerAlaThrVal 662
Qy      343 AACTTTTCGACATCAAGAAAGAGTTGAATCAAAACTTGGAATTCATTAACAATAC 402
Db      :|||      :|||      :|||      :|||      :|||
663 AenAenThrAaSerAlaAaSerGln-----AlaLySgInGlnAlaIleAaThrValAa 679
Qy      403 AATATGACTCAAAAGCAACCACTTATCTTAATTAATTAAGAAAGAGAGCGCA 462
Db      |||      :|||      :|||      :|||      :|||
680 ThrAaAaAa-----GlnAenIleIleSerLyAaAaenGlyAaenAaIleAaThrGln 696
Qy      463 GCAGTACCTTGAAGAACCACTTACGCTCAAGTTCTAAAGGTAATTCGAAAAAT 522
Db      :|||      :|||      :|||      :|||      :|||
697 AlaGlu-----ValGlnAlaIleIleLySgInValAaenAaIleAaLyS 710
Qy      523 GACCACTCAATCAAAATAGCAAGTGTGTTGGCTGTGTAAGGCAAGCGGCTTCCCT 582
Db      |||      :|||      :|||      :|||      :|||
711 GlnAlaLeuAaenGlyAaenAaAaen----- 718
Qy      583 TTGAAAGGCAATGATTAAGTTGATCTCAATCAATGAGTACGCTTCAAGATCAAGAA 642
Db      :|||      :|||      :|||      :|||      :|||
719 ---ValGlnIleAaIleAaSerLyAaSerAlaIleAaenAaSerAaenAaen 737
Qy      643 TTGGCTCAAGAAATGACAATCTCAATCAAGCGGTA 678
Db      |||      :|||      :|||      :|||      :|||
738 GlnAlaGlnLySgInLySgInLySgInGlnVal 748

```

```

RESULT 13
US-10-369-493-3209
/ Sequence 3209, Application US/10369493
/ Publication No. US2003023675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 3209
/ LENGTH: 1361
/ TYPE: PRT
/ ORGANISM: Neurospora crassa
US-10-369-493-3209

```

```

Alignment Scores:
Pred. No.: 0.0043      Length: 1361
Score: 127.00      Matches: 76
Percent Similarity: 37.2%      Conserved: 40
Best Local Similarity: 24.4%      Mismatches: 92
Query Match: 10.8%      Indels: 104
DB: 4      Gaps: 14

```

US-09-360-685C-26 (1-685) x US-10-369-493-3209 (1-1361)

```

Qy      16 AAGGATTCAGACAGTAAACGCAACCAAAAGCGAC--CTTGAATAATCCGTTAAAGAT 72
Db      |||      :|||      :|||      :|||      :|||
1003 LyAaSerLyLeuLySgInSerGlnThrGlnAaSerAlaLeuAaenAaenAaIleLyAaSer 1022
Qy      73 GTC-----ATCATCAAT 84
Db      |||      :|||      :|||      :|||      :|||
1023 MetLySerThrLySgInLySgInLySgInAaSerAlaLySgInThrValIleuAa 1042
Qy      85 CAAGC-----GTACGGAATTAAGTTGATTAATCTCAATCAAGCGGTA 126
Db      :|||      :|||      :|||      :|||      :|||
1043 GlnLySgInMetLySgInIleGlnLySgInLySgInAaSerIleAaenAaGlnAaenGlnAa 1062
Qy      127 TCAATGCGCTTAAGCAACGCGTGAATTCAGTAGG-----GTAGACCAAGCTTACCGAT 180
Db      :|||      :|||      :|||      :|||      :|||
1063 SerThrLyAaenAaIleThrLeuAaSerLyAaSerGlnIleIleAaSerGlnLeuAaSer 1082
Qy      181 CTCAA----- 186
Db      |||      :|||      :|||      :|||      :|||
1083 IleLyThrAlaAaenSerThrIleAaSerThrLeuAaGlyAaSerValLyAaSerLyAaSerAla 1102
Qy      187 -----AATTTCTCAAGAGCAATGCGCCCAACAGCTCA----- 222
Db      :|||      :|||      :|||      :|||      :|||
1103 IleLeuAaIleAaSerThrLyAaSerValAlaIleAaSerAlaGlnIleuAaIleLySgIn 1122
Qy      223 -----AAAAATGAAGCTCAATGCTTAAGAAAAAATCTGAATATATCA 267
Db      :|||      :|||      :|||      :|||      :|||
1123 AlaGlnIleAaSerLyAaSerAaIleAaLeu-----AlaLySgInGlnIleuAaIleAa 1141
Qy      268 TCCGTTAAGATGTGTGAATGGAACCTGATGCGTAATGCTTATCTCAAGCAAGGCC 327
Db      :|||      :|||      :|||      :|||      :|||
1142 PheGlnLySgInValGlnThrLeuThrAaSerGlnAlaLySgInLeuAaenAaSerValAla 1161
Qy      328 ACA-----ACTCTTTCTAAAACTTTCCGACATC 357
Db      :|||      :|||      :|||      :|||      :|||
1162 ThrLySerThrGlnLeuAaGlnAaSerAlaThrIleSerLySgInLySgInLySgIn 1181
Qy      358 -----AGAAAGAGTTGAATCAAACTTGGAAT 387
Db      :|||      :|||      :|||      :|||      :|||
1182 PheAaSerLySerThrAaSerValThrLySgInLySgInGlnIleuAaenSerThrLyAaSerAlaAaen 1201
Qy      388 TTCAT-----AACATTAACAATATGACTCAAA-----AACGA 423
Db      :|||      :|||      :|||      :|||      :|||
1202 LeuThrGlnLyAaGlyGlnIleGlySerAaSerAaIleGlyLeuAaIleAaLySgIn 1221
Qy      424 CCATTTATGCTTAA-----GTATTAATAAAGAAAGCAAGCAAGCAAGCTTGA 477
Db      :|||      :|||      :|||      :|||      :|||
1222 GlnLeuAaSerGlnAaGlnAaIleAaIleAaIleAaIleAaIleAaIleAaIleAaIleAa 1241
Qy      478 GAACCACTTACGCTCAAGTTGCTTAAGAAAGTAATCAAAATTCAG-----CGA 528
Db      :|||      :|||      :|||      :|||      :|||
1242 LyAaenValLySgInLeuThrAaSerGlnAaIleAaIleAaIleAaIleAaIleAaIleAa 1261
Qy      529 CTCATCAATATGCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTTCCCTTGA 588
Db      :|||      :|||      :|||      :|||      :|||
1262 AaPThrGlnLeuAaIle-----Gln 1267
Qy      589 AGCAGATGATTAAGTATGATCTCAATGATGAGTACGCTTCAAGGAATCAAGAAATTCGCT 648
Db      :|||      :|||      :|||      :|||      :|||
1268 AaPlyAaSerAlaIleSerLySgInLySgInAaSerIleAaIleAaLySgInGlnIleuAa 1287
Qy      649 CAGAAATTCATCAATCTCAATCAAGCGGTAATCAAGAA 664
Db      |||      :|||      :|||      :|||      :|||
1288 ThrLyAaSerAlaSerLeuThrGlnLySgInThrGln 1299

```

```

RESULT 14
US-09-815-242-5835
/ Sequence 5835, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsson, Karl L.
/ APPLICANT: Zykking, Judith W.
/ APPLICANT: Wall, Daniel

```


Tue Mar 7 12:58:25 2006

us-09-360-685c-26.rapbm

Page 12

[illegible]

Search completed: March 6, 2006, 20:22:24
Job time : 176.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 6, 2006, 20:00:26 ; Search time 7.1 Seconds
(without alignments)
1595.291 Million cell updates/sec

Title: US-09-360-685C-26
Perfect score: 1171
Sequence: 1 aaaaatgcaataaga.....caatcaagcgatcagaag 685

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues
Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-Q=/abs/ABSWEB.spool/US0936085/runat_06032006_121320_15444/app_query.fasta_1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-HOST=abgs02p -USER=US0936085@cgn1_1.71@runat_06032006_121320_15444
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*

- 1: /cgn2_6/prodata/1/1aa/5-COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/6-COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/H-COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/CTUS-COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/RE-COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/backfill1est.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	96.2	1147	2	US-08-470-260-5 Sequence 5, Appli
2	1127	96.2	1147	2	US-08-471-491-5 Sequence 5, Appli
3	1127	96.2	1147	2	US-08-466-662-5 Sequence 5, Appli
4	1127	96.2	1147	2	US-08-477-451-2 Sequence 2, Appli
5	1081	92.3	3200	1	US-08-477-451-8 Sequence 8, Appli
6	1030.5	88.0	268	2	US-08-993-010-4 Sequence 4, Appli
7	949	81.0	3287	1	US-08-477-451-7 Sequence 7, Appli
8	930	79.4	3169	1	US-08-477-451-6 Sequence 6, Appli
9	927.5	79.2	3174	1	US-08-477-451-3 Sequence 3, Appli
10	902.5	77.1	1181	1	US-08-053-614-4 Sequence 4, Appli
11	902.5	77.1	1181	1	US-08-316-397B-4 Sequence 4, Appli
12	902.5	77.1	1181	1	US-09-034-306-4 Sequence 4, Appli

13	902.5	77.1	1181	2	US-09-259-437-4 Sequence 4, Appli
14	902.5	77.1	1181	4	PCT-US93-09782-4 Sequence 4, Appli
15	798.5	68.2	3177	1	US-08-477-451-4 Sequence 4, Appli
16	454.5	38.8	859	1	US-08-053-614-2 Sequence 2, Appli
17	454.5	38.8	859	1	US-08-316-397B-2 Sequence 2, Appli
18	454.5	38.8	859	1	US-09-034-306-2 Sequence 2, Appli
19	454.5	38.8	859	2	US-09-259-437-2 Sequence 2, Appli
20	454.5	38.8	859	4	PCT-US93-09782-2 Sequence 2, Appli
21	123	10.5	10182	2	US-09-134-001C-3159 Sequence 3159, Ap
22	120.5	10.3	5024	2	US-09-710-279-2564 Sequence 2964, Ap
23	118	10.1	1693	2	US-09-560-385A-4 Sequence 4, Appli
24	118	10.1	1693	2	US-09-560-385A-8 Sequence 8, Appli
25	118	10.1	1713	2	US-08-600-982-24 Sequence 24, Appli
26	118	10.1	1713	2	US-09-560-385A-6 Sequence 6, Appli
27	118	10.1	1713	2	US-09-538-092-1359 Sequence 1359, Ap
28	118	10.1	1713	4	PCT-US94-10261A-24 Sequence 24, Appli
29	118	10.1	1724	2	US-09-560-385A-2 Sequence 2, Appli
30	117	10.0	693	2	US-08-235-836C-68 Sequence 68, Appli
31	115.5	9.9	1095	2	US-09-710-279-3154 Sequence 3154, Ap
32	115	9.8	873	2	US-09-336-447A-13 Sequence 13, Appli
33	115	9.8	873	2	US-09-952-267B-13 Sequence 13, Appli
34	114	9.7	3712	2	US-10-037-417-48 Sequence 48, Appli
35	113.5	9.7	1401	2	US-09-750-590A-2 Sequence 2, Appli
36	113.5	9.7	1939	2	US-09-538-092-917 Sequence 917, App
37	113.5	9.7	1939	2	US-09-538-092-917 Sequence 917, App
38	113	9.6	878	2	US-09-540-236-3401 Sequence 3401, Ap
39	113	9.6	3696	2	US-09-134-001C-5080 Sequence 5080, Ap
40	112.5	9.6	611	2	US-09-710-279-2586 Sequence 2586, Ap
41	112.5	9.6	718	2	US-09-540-236-2753 Sequence 2753, Ap
42	112	9.6	1935	2	US-09-538-092-916 Sequence 916, App
43	112	9.6	3712	2	US-10-037-417-51 Sequence 51, Appli
44	111.5	9.5	700	1	US-07-720-588-2 Sequence 2, Appli
45	111.5	9.5	700	1	US-08-785-190-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-470-260-5
Sequence 5, Application US/08470260
Patent No. 6077706
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
for Vaccines and Diagnostics
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001

QY 661 AATTCATCAAGCGGTATCAGAA 684
DB 970 AenLeuAenGlnAlaValSerGlu 977

RESULT 3
US-08-466-662-5
Sequence 5, Application US/08466662B
Patent No. 6130059
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
TITLE OF INVENTION: Diagnostics
FILE REFERENCE: CHIR0057
CURRENT APPLICATION NUMBER: US/08/466,662B
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRF
ORGANISM: Helicobacter pylori
US-08-466-662-5

Alignment Scores:
Pred. No.: 5,53e-108 Length: 1147
Score: 1127.00 Matches: 228
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.2% Indels: 0
Gaps: 0

US-09-360-685c-26 (1-685) x US-08-466-662-5 (1-1147)

QY 1 AAAAAAGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAT 60
DB 750 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerAenGluAen 769
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGTAAAGTAAAGTAAATCATCATCA 120
DB 770 SerValLysAenValIleIleAenGlnLysValThrAenLysValAenAenAenGln 789
QY 121 GCGGTATCAGTGGCTTAAGCAAGCGGTGAATTCAGTAGGTAGAGCAAGCTTACCCGAT 180
DB 790 AlaValSerValAlaLysAlaThrGlyAenPheSerArgValGluGlnAlaLysAen 809
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCTCAAT 240
DB 810 LeuLysAenPheSerLysGlnLysAenGlnGlnAlaGlnLysAenGlnLysAen 829
QY 241 GCTAAGAAAAATCTGAATATATCAATCCGTAAAGTGTGAATGCAACCTTATGC 300
DB 830 AlaAenGlyLysSerGlnLysValIleValSerValLysAenGlyValAenGlyThrLeuVal 849
QY 301 GGTAAATGGTATCTCAAGCAAGCAAGCAAGCTTCTTAAAACTTTTGGACATCAAG 360
DB 850 GlyAenGlyLysSerGlnAlaGlnLysValThrLysSerLysAenPheSerAenLys 869
QY 361 AAGAGTTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGACTCAAAAAC 420
DB 870 LysGlnLysAenAlaLysLysGlnLysAenPheAenAenAenAenAenGlyLysLysAen 889
QY 421 GAACCAATTATGTAAGTAAATTAAGAAAGCAAGCGCAAGCTAGCCTTGAAGAA 480
DB 890 GluPheLysValLysValAenLysLysValAenGlyGlnAlaLysSerLeuGln 909
QY 481 CCCATTATCGCTCAAGTGTCAAAAAAGTAAATGCAAAATGACCGACTCAATCAATA 540
DB 910 ProLysValLysGlnValAlaLysLysValAenAlaLysLysLysAenPheLysAenGln 929

QY 541 GCAAGTGTGGTGGTGTGAGGCAAGCGGCTTCCTTGAAGGCAATCAATAA 600
DB 930 AlaSerGlyLysGlyValValGlyGlnAlaLysGlyPheProLeuLysArgHisAenLys 949
QY 601 GTTGATGATCTCAGTAAGGAGGCTTCAAGAAATCAAGAAATGGCTCAGAAATTCAC 660
DB 950 ValAenAenLysSerLysValGlyLysSerAenGlnLysAenGlnLysLysAen 969

QY 661 AATTCATCAAGCGGTATCAGAA 684
DB 970 AenLeuAenGlnAlaValSerGlu 977

RESULT 4
US-08-477-451-2
Sequence 2, Application US/08477451
Patent No. 592865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSER: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335, 002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-2

Alignment Scores:
Pred. No.: 7.47e-108 Length: 3289
Score: 1127.00 Matches: 228
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.2% Indels: 0
Gaps: 0

US-09-360-685c-26 (1-685) x US-08-477-451-2 (1-3289)

QY 1 AAAAAAGCAAAATTAAGATTTCAGCAAGTAAAGCAAAAGCAAGCACTTGAAT 60
DB 2281 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerAenLysAen 2300
QY 61 TCCGTTAAAGATGATCATCAATCAAAAGTAAAGTAAAGTAAATTCATCAATCA 120
DB 2301 SerValLysAenValIleIleAenGlnLysValThrAenLysValAenAenAenGln 2320
QY 121 GCGGTATCAGTGGCTTAAGCAAGCGGTGAATTCAGTAGGTAGAGCAAGCTTACCCGAT 180

Tue Mar 7 12:58:24 2006

US-09-360-685C-26.ra1

Page 4

Db 2321 AlaValSerValAlaIySaIaThrGlyAspPheSerArgValIGluGlnAlaLeuAlaAsp 2340
QY 181 CTCAAAAAATTTCTCAAAAGGACGATTTGGCCCAACGCTCAAAAAATTAAGCTCAAT 240
Db 2341 LeuIySaenPheSerIySGluGlnLeuAlaGlnGlnAlaGlnIySaenGlnSerLeuSaen 2360
QY 241 GCTAGAAAAAATCTGAATATATATCAATCCGTTAAGAAATGCTGAATGCAACCTTATG 300
Db 2361 AlaArgIySaenSerGluIleIyGlnSerValIySaenGlnIyValaGlnIyThLeuVal 2380
QY 301 GGTAAATGGTTATCTCAAGCAAGAACCAACCTTTCTTAAATCTTTCCAGATCAAG 360
Db 2381 GlnaenGlyLeuSerGlnAlaGlnAlaThrThrLeuSerIySaenPheSerAspIleIyS 2400
QY 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAATGAATCAAAAC 420
Db 2401 IySGluLeuSaenAlaIyLeuGlnIySaenPheSaenSaenSaenSaenSaenIyLeuIySaen 2420
QY 421 GAACCAATTAATGCTAAAGCTTAATTAAGAAAGCAAGGCAAGCACTAGCTTGAAGAA 480
Db 2421 GluProIleIyThIaIySaenIySaenIySaenIySaenIySaenIySaenIySaen 2440
QY 481 CCCATTAAGCTCAAGCTTGTAAATTAATGCAAAATTAATGCAAGCACTCAATCAATA 540
Db 2441 ProIleIyThIaIySaenIySaenIySaenIySaenIySaenIySaenIySaen 2460
QY 541 GCAAGCTGTTGGGCTTTGAAGGCAAGCAAGGCTTCCCTTGAAGCAAGCAATGAATA 600
Db 2461 AlaSerGlyLeuGlnIyValIyGlnIyAlaIySaenIySaenIySaenIySaen 2480
QY 601 GTTGAATGATCTCAAGTAAGGAGCTTTCAAGCAATCAAGATTTGGCTCAAGAAATTCAG 660
Db 2481 ValAspAspLeuSerIySaenIyLeuSerArgSaenGlnIyLeuAlaGlnIyIleAsp 2500
QY 661 AATCTCAATCAAGCGGTATCAAGA 684
Db 2501 AsnLeuSaenGlnAlaIyValIySerGln 2508

RESULT 5
US-08-477-451-8
Sequence 8, Application US/08477451
Patent No. 592865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-655-3542
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3200 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-8

Alignment Scores:
Pred. No.: 4,22e-103
Score: 1081.00
Percent Similarity: 97.3%
Best Local Similarity: 97.3%
Query Match: 92.3%
DB: 1
Length: 3200
Matches: 220
Conservative: 0
Mismatches: 0
Indels: 6
Gaps: 6

US-09-360-685C-26 (1-685) x US-08-477-451-8 (1-3200)

QY 678 TACCGCTGATTTGAGATTTGCAATTTCTGAGCAATTTCTGATTCCTGAAGCCCTTAC 619
Db 768 TyArgLeuIleGlnIleValAlaAsnPheLeuSerGlnPheLeuIlePro---LysProIy 786
QY 618 CTYACGAGATCAATCAACTTATATCATGCTTTTCAAGGAAAGCCCGCTGCTCCCTAC 559
Db 787 LeuThrGlnIleIleAsnPheIleMetProPheGlnArgGlnAlaArgCysLeuProIy 806
QY 558 AACACCCAAACCACTGCTATTTGATTTGATGTCGTCGCAATTTTTCATTTACCTTTTAC 499
Db 807 AsnThrGlnThrCysArgLeuIleGlnSerValaAsnPheCysIleIyThLeuPheSer 826
QY 498 AACTGAGCGTAAATGGGTTCTTCAAGGCTACCTTCCCTGCTTCTTTTATTAAC 439
Db 827 AsnLeuSerValaAsnGlyPhePheIySaIaSerCysLeuProCysPheLeuPheIleAsn 846
QY 438 TTGAGCAATTAATGGGTTGCTTTGAGTCCATTTATTTGTTATTTGAATTTTCAG 379
Db 847 PheSerIleAsnGlyPheValPheGlnSerIleIleValIleGlnIleSerIyS 866
QY 378 TTGAGCAATCACTTTCTTGAATGTCGAAAGTTTGAAGAGTTGGCTTCTGC 319
Db 867 PheCysIleGlnLeuPheLeuAspValaArgIySaIaPheArgIySerCysGlyPheCys 886
QY 318 TTGAGTAACCAATTAACGACGACGATTCATTCACCATTTCTTAACGATTTGATATAT 259
Db 887 LeuArg---ProIleThAsp---GlySerIleIleThrIleLeuAsnGlyLeuIleIy 904
QY 258 TTGAGATTTTCTGACATTTGAGATTTCTCAATTTTGTAGCTTTGGCCCAATTTGCT 139
Db 905 PheArgPhePheSerSerIleGlnThrPheIlePheLeuSerLeuLeuGlnIleLeu 924
QY 138 CTTTGAGAAATTTTGAATCGGCTTACGCTTCAACCCCTGAAATCAACCCGTTGC 139
Db 925 Leu---GlnIlePheGlnIleGly---ArgLeuLeuIyProThrGlnIleThrArgCys 942
QY 138 TTGAGCACTGATNACGCTTGAATGATTAATCACTTATTCGTTTACCTTTGATGAT 79
Db 943 PheSerHis---TyArgLeuIleGlnIleIleAsnPheIleArgIyLeuLeuIleAsp 961
QY 78 GATCACTCTTTAAGCAATTTTCAAGTGGCTTTTCTGCTTCTGCTTACCTTGAATTC 19
Db 962 AspHisIlePheAsnGlyIlePheIyValaIaPheCysLeuArgIyLeuIleGlnIle 981
QY 18 CTYATTTTGGCAATTTT 1
Db 982 LeuIlePheAlaIlePhe 987

RESULT 6
US-08-993-010-4
Sequence 4, Application US/08993010
Patent No. 6902903
GENERAL INFORMATION:
APPLICANT: Quan, Stella
APPLICANT: Valenzuela, Pablo
TITLE OF INVENTION: HELICOBACTER PYLORI DIAGNOSTICS
FILE REFERENCE: 1360,002

```

:
: CURRENT APPLICATION NUMBER: US/08/993,010
:
: CURRENT FILING DATE: 1997-12-18
:
: EARLIER APPLICATION NUMBER: 60/033,707
:
: EARLIER FILING DATE: 1996-12-19
:
: NUMBER OF SEQ ID NOS: 4
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 4
:
:   LENGTH: 268
:   TYPE: PRT
:   ORGANISM: Helicobacter pylori
:
: US-08-993-010-4

```

Alignment Scores:	
Pred. No.:	3,45e-98
Score:	1030.50
Percent Similarity:	83.4%
Best Local Similarity:	81.9%
Query Match:	88.0%
DB:	2
Length:	266
Matches:	21
Conservative:	4
Mismatches:	7
Indels:	3
Gaps:	37

US-09-360-685C-26 (1-685) X US-08-993-010-4 (1-268)

Qy	1	AAAAATGGCAAAAATAAGATTTCAGCAAGTAAAGCAAGCAAAAAGGACCTTGAAT	60
Db	4	LyAsnGcLybAsnInyAspPheSerIysValIntrGlnAlaYsSerAspLeuGlnSn	23
Qy	61	TCGCTTAAAGATGTGATCATCAATCAAAAGCTAAAGCATTAAGTTGATTAATCTCAATCAA	120
Db	24	SeTlleYAspValIllePheAsnGlnYsIlleThrAspIysValAspAspLeuAsnGln	43
Qy	121	GGCGTATCAATGGCTTAAAGCAACGGGTGATTTCAAGTAGGTAGACCAAGCCTTAGCCGAT	180
Db	44	AlaValSerValAlaIaYsAlaThrTrgIAspPheSerAspValGluGlnAlaLeuAlaAsp	63
Qy	181	CTCAAAAATTTCTCAAAAGCAAAATGGCCCAACAGCTCAAAAATAAGAAAGTCTCAAT	240
Db	64	LeuIlyAsnPheserIysGluGlnIleuAlaGlnGlnAlaGlnIlyAsnGlnSerLeuAsn	83
Qy	241	GCTAGAAAAAATCTGAAATATATATCAATCCGTTTAAAGATGGTGTAATGGAACCTTAGTC	300
Db	84	AlaGcLybYsSerGlnIleTrgIAsnSerValIyAsnGcIyAlaAsnGcIyThrLeuVal	103
Qy	301	GGTAATGGGTATCTCAAGCAGAGCCACACACTCTTTCTTAAAACTTTTGGACATCAAG	360
Db	104	GlyAsnGcIyLeuSerGlnAlaGlnAlaThrThrLeuSerIyAsnPheSerAspIleYs	123
Qy	361	AAAGATTTGAATGCAAAATTT---GGAATTTGCATTAACAATTAACAATATGACTCAA	417
Db	124	LybGcIyLeuAsnAlaIyLeuPheGcIyAsnPheAsnAsnAsnAsnAsnGcIyLeuIyS	143
Qy	418	AAC-----GAACCCATTATGCTTAAAGTTAATAAAAAAGAAAGCAAGGGCAAGAGCTTAC	471
Db	144	AsnSerThrGluProIleTrgIAlaIyValAsnIySlySlyThrGcIyGlnValAlaSer	163
Qy	472	CTTGAAGAAACCATTTAGCGCTCAAGTTGCTAAAAAGTAAATGCAAAAATTGACCGAATC	531
Db	164	ProGluGlnProIleTrgIThrGlnValAlaIySlyValAsnAlaIySilleAspArgLeu	183
Qy	532	AATCAAAATGCAAGTGGTTGGGTGTGTAAGGGCAAGCAAGCGGGC-----	576
Db	184	AsnGlnIleAlaSerGcIyLeuGcIyGlnIyValGcIyValAlaAlaGcIyPheProLeuIySArg	203
Qy	576	-----	576
Db	204	HisAspIyValAspAspLeuSerIySValGcIyAlaGserValSerProGluProIleTrg	223
Qy	577	-----TTCCCTTGAAGAGCATGATTAAGTTGATGAT	609
Db	224	AlaThrIleAspAspLeuGcIyGcIyProPheProLeuIySArgHisAspIySValAspAsp	243
Qy	610	CTCAGTAGGTAGGGGCTTTCAAGGAATCAAGAAATGGCTCAAGAAATTTGCAATTCAT	669
Db	244	LeuSerIyValGcIyLeuSerArgAsnGlnIleuAlaGlnIySilleAspAsnLeuAsn	263

QY	670	CAAGCGTATCAGAA	684
Db	264	GlAlaValSerGlu	268

RESULT 7
US-08-477-451-7
; Sequence 7, Application US/08477451

GENERAL INFORMATION:

;	TITLE OF INVENTION:	Helicobacter
:	NUMBER OF SEQUENCES:	46

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation

STREET: 4360 HOLLOMAN
CITY: Emeryville

COUNTRY: USA

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS/MS DOS

```

; SOFTWARE: PatentIn Re
; CURRENT APPLICATION DATA

```

APPLI CATION NUMBER: U
FILING DATE: 07-JUN-1

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION

REGISTRATION NUMBER:

TELECOMMUNICATION INFORMATION
FEBRUARY 1983 201-27TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO.:

SEQUENCE CHARACTERISTICS

```

; TYPE: amino acid
; STRANDEDNESS: single

```

MOLECULE TYPE: protein

0
0
0
0
0
0
0
0

Pred. No.:	1.88e
------------	-------

Percent Similarity:	94.3%
---------------------	-------

```
Query Match: 81.0%
DR: 1
```

US-09-360-685C-26 (1-685) x US

QY 683 TCTGATACCGCTTGA

Db 787 SerAspThrAla---1

QY 623 CCTACCTTACTGAGAT
|||||

Db 804 ProThrLeuLeuArg;

563 CCATCAATCCCAAA
QY

023 FLORENCE

100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Alignment Scores:

Pred. No.:	1,88e-89	Length:	328
Score:	949.00	Matches:	214
Percent Similarity:	94.3%	Conservative:	0
Best Local Similarity:	94.3%	Mismatches:	1
Query Match:	81.0%	Indels:	12
DB:	1	Gaps:	11

US-09-360-685C-26 (1-685) x US-08-477-451-7 (1-3287)

[illegible]

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.rtf

Page 6

Db	861	LeuThrIeuAla---MetGlySerPheIeuSerProIeuLeuLeuLeuLeuIuysPhe	879
Qy	383	CGAAGTTCGACATCAACTCTTCTTGAGATGTCGCAAAAGTTTTCAGAAAGATGTGGCT	324
Db	880	ProSerPhehAlaPheAlaSerPheIeuIeuIeuSerIuyluPheIeuGluValValAla	889
Qy	323	TCCTCTGGAGATACCCATTCACGAAATGGAGTGGCTTCATCAACCATTCCTTAACGANTGA	264
Db	900	SerAla---AspAsnProIeuProIuAlaGValProIeuIuProPheIeuIuIuAsp---	917
Qy	263	TATATTCAGATTTTTTCTTCAGATGATGAGATTCATTTTTTTAGCTTGTGGCCAAAT	204
Db	918	TyrIleSerAspPhePheIeuAlaIeuAluIuIuIeuSerPhePhe--AlaCysTrpAlaAsn	936
Qy	203	TGTCCTTTGAGAAATTTTGGATGATGACGCTTACGCTTGCTGACCTTACTGAATCAACC	144
Db	937	CysSerPheGluIuyluPheIeuAluIuIuSerIuAlaAlaCysSerTruIeuIeuIuysSerPro	956
Qy	143	GTGGCTTTTGGCAACGATTAACCGCTTGATGTGAGATTAACAATTCATTCCTTACCTTTTGA	84
Db	957	ValAlaIeuAlaIuIuIuIuSerTruAla---LeuArgIeuSerTruIeuSerValIuIuPhe---	974
Qy	83	TTGATGATACATCTTTTAACGAAATTTTCAAGGGCGCTTTTGGTTCCTGTTACTGTGCTG	24
Db	975	IeuIeuIleIuIuIuSerIeuIuIuIuIuPheSerIuSerIeuPheAlaCysValTruIeuIeu	994
Qy	23	AAATCCTTATTTTGGCATTT 3	
Db	995	IuylSerIeuIuIuIeuProPhe 1001	

```

? RESULT 8
? US-08-477-451-6
? Sequence 6, Application US/08477451
? Patent No. 5928865
? GENERAL INFORMATION:
? APPLICANT: Covacci, Antonello
? TITLE OF INVENTION: 'Helicobacter Pylori CagI Region
? NUMBER OF SEQUENCES: 46
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Chiron Corporation
? STREET: 4560 Horton Street
? CITY: Emeryville
? STATE: CA
? COUNTRY: USA
? ZIP: 94608-2916
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/477,451
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: McClung, Barbara G.
? REGISTRATION NUMBER: 33,113
? REFERENCE/DOCKET NUMBER: 0335.002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 510-601-2708
? TELEFAX: 510-655-3542
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3169 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-477-451-6

```

Alignment Scores:	
Pred. No.:	1,71e-87
Score:	930.00
Length:	3169
Matches:	205

Percent Similarity:	89.9%	Conservative:	0
Best Local Similarity:	89.9%	Mismatches:	1
Query Match:	79.4%	Indels:	22
DB:	1	Gaps:	20

[illegible]

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,451
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: McClung, Barbara G.
 REGISTRATION NUMBER: 33,113
 REFERENCE/DOCKET NUMBER: 0335.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 510-601-2708
 TELEFAX: 510-655-3542
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3174 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-451-3

Alignment Scores:
 Pred. No.: 3,116-87 Length: 3174
 Score: 927.50 Matches: 215
 Percent Similarity: 94.3% Conservative: 0
 Best Local Similarity: 94.3% Mismatches: 0
 Query Match: 79.2% Indels: 13
 DB: Gaps: 13

US-09-360-685C-26 (1-685) x US-08-477-451-3 (1-3174)

```

QY 2 AAAATGGCAAAATTAAGATTTCAGCAAGTAACGCAAGCAAAACCACTTGAAATT 61
DB 2201 LysMetAlaLysIleArgIleSerAlaArg---ArgLysGlnLysLethrLeuLysIle 2219
QY 62 CCGTTAAAGATGTGATCATCATCAATCAAAAGGTAAACGGTAAAGTGAATATCTCAATCAG 121
DB 2220 ProLeuLysMet---SerSerIleLysArg---ArgIleLysLeuIleSerIleLys 2237
QY 122 CGGATCAAGTGGCTAAAGCAACGGGTGATTTCAGTAGGAGTAAAGCAACGGTAAAGCGATC 181
DB 2238 ArgLysGlnTrpLeuLysGlnArgValIleSerValGly---SerLysArg---ProIle 2255
QY 182 TCAAAATTTCTCAAAAGCAAGCAATTCGCCCAACACCTCAAAATAATGAAGTTCATG 241
DB 2256 SerLysIleSerGlnArgSerAsnTrpProAsnLysLeuLysLysMetLysValSerMet 2275
QY 242 CTAGAAAAAATCTGAAATATATATCATCGTTAAGAAAGTGTGAATGAACCTAGTCG 301
DB 2276 LeuGlnLysAsnLeuLysTrpIleAsnProLeuArgMetVal---MetGlnPro---Ser 2293
QY 302 GTAATGGGTTATCTCAAGCAAGCAACCACTCTTTCTAAAACCTTTGACATCAAGA 361
DB 2294 ValMetGlyTrpLeuLysGlnLysProGlnLeuPheLeuLysThrPheArgThrSerArg 2313
QY 362 AAGAGTTGAATGCAAACTTGAAATTTCAATPAACAATAACATAATGACTCAAAAAGC 421
DB 2314 LysSer---MetGlnAsnLeuGlnIleSerIleThrIleMetAspSerLysTrp 2332
QY 422 AACCATTTAAGTAAAGTAAATAAAGCAAGCGGCAAGCAAGCAAGTACCTTGAAGAAC 481
DB 2333 AsnProPheMetLeuLysLeuLysArgLysGlnGlyLysGlnLeuLysLeuLysAsn 2352
QY 482 CCATTACGCTCAAGTGTCTAAAAAAGTAAATGCAAAAATGACGCACTCAATCAAAATAG 541
DB 2353 ProPheThrLeuLysLeuLysArg---MetGlnLysLeuThrAspSerIleLys--- 2370
QY 542 CAAGTGTGTTGGGTGTGTAGGCAAGCAAGCGGCGCTTCCCTTTGAAAAGCAATGAATAG 601
DB 2371 GlnValIleAlaTrpValLeu---GlyLysGlnArgAlaSerLeu---LysGlnMetIleLys 2388
QY 602 TTGATGATCTCAGTAAGGTAGGCTTTCAAGGAATCAAGAATTGGCTCAGAAAATTGACA 661
  
```

RESULT 10
 US-08-053-614-4
 Sequence 4, Application US/08053614
 Patent No. 5403924
 GENERAL INFORMATION:
 APPLICANT: COVER, TIMOTHY L.
 APPLICANT: BLASER, MARTIN J.
 TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
 TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 STREET: 133 Carnegie Way, Suite 400
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/053,614
 FILING DATE: 19930426
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: SPRATT, GWENDOLYN D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.009
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1181 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-053-614-4

Alignment Scores:
 Pred. No.: 96-85 Length: 1181
 Score: 902.50 Matches: 195
 Percent Similarity: 76.1% Conservative: 6
 Best Local Similarity: 73.9% Mismatches: 26
 Query Match: 77.1% Indels: 37
 DB: Gaps: 4

US-09-360-685C-26 (1-685) x US-08-053-614-4 (1-1181)

```

QY 1 AAAATGGCAAAATTAAGATTTCAGCAAGTAACGCAAGCAAAACCACTTGAAATT 60
DB 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspGlnGlnAsn 769
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAACGGTAAAGTGAATCTCAATCAA 120
DB 770 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspGlnLeuAsnGln 789
QY 121 GCGGTATCAGTGGCTAAAGCAACGCGGTGATTTCAGTAGGAGTAAAGCAAGCTTACCGCAT 180
DB 790 AlaValSerValAlaLysIleAlaCysAspPheSerGlyValGlnGlnAlaLeuAlaAsp 809
QY 181 CTCAAAAATTTCTCAAGGAGCAATTCGCCCAACAGCTCAAAAATAATGAAGTCTCAAT 240
  
```

Db 810 LeuLyAsenPheSerLySGluGlnLeuAlaGlnGlnAlaGlnLyAsenGluSerPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCATCGCTTAAGAGTGAGTGAATGACCCATGCTC 300
Db 830 Val---GlyLySerGluLeuIleTyrGlnSerValIleAsnGlyAlaAsnGlyThrLeuVal 848
QY 301 GGTAAAGGTGATCTCTCAAGAGACGCAACTCTTCAAAAGCTTTTGGACATCAAG 360
Db 849 GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLyAsenPheSerPheIleAsn 868
QY 361 AAAGAGTTGAATGCAAACTGGAAATTTCAATTAACATTAATGACATCAAAAC 420
Db 869 LySGluLeuAsnGluIlySphelyAsnPhenAsnAsnAsnAsnAsnGlyLeuLyAsn 888
QY 421 -----GAACCATTTATGTAAAGTTATTAATTAAGAAAGGCAAGGCAAGCTAGCCTT 474
Db 889 GlyGlyGluProIleTyrAlaGlnValAsnLySphThrGlyGlnValAlaSerPro 908
QY 475 GAAGAACCATTTACGCTCAAGTCTTAAAGGTAATGCAAAATTTGACGACTCAAT 534
Db 909 GluGluProIleTyrAlaGlnValAlaLySphValThrLySphIleAsnGlnLeuAsn 928
QY 535 CAATTAAGCA---AGTGGTTGGCTGTGAGGCAAGCA----- 570
Db 929 GlnAlaAlaThrSerGlyPheGlyValGlyGlnAlaGlyPheProLeuLySphArgHis 948
QY 570 ----- 570
Db 949 AspLySphValGluAspLeuSerLySphValGlyArgSerValSerProGluProIleTyrAla 968
QY 571 -----GCGGAGCTTCCTCTTGAAGAGCAATGAATGATGATGCTC 612
Db 969 ThrIleAspAspLeuGlyGlySerPheProLeuLySphArgHisAspLySphValAspAspLeu 988
QY 613 AGTAAGGTAGGGCTTTCAAGAAATCAAGAAATGGCTCAAGAAATTTGACATCTCAAA 672
Db 989 SerLySphValGlyLeuSerArgAsnGlnGluLeuThrGlnIlySphIleAsnLeuSerGln 1008
QY 673 GCGGTATCAAGAA 684
Db 1009 AlaValSerGln 1012

RESULT 11
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOS
; APPLICANT: TUMURU, MURALI K. R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P. C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-4
Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 1 Gaps: 4
US-09-360-685c-26 (1-685) x US-08-316-397B-4 (1-1181)
QY 1 AAAAAATGCAAAATTAAGATTTCAAGAGTAAAGCAAGCAAAAGCACTTGAAAT 60
Db 750 LyAsnGlyLySphAspPheSerLySphValThrGlnAlaLySphSerPheGlnGluAsn 769
QY 61 TCCGTTAAAGATGTGATCATCATCAAAAGCTAAGGTAAGTGAATGATATCTCAATCA 120
Db 770 SerIleLySphValIleIleAsnGlnIlySphThrPheLySphValAlaSerPro 789
QY 121 GCGGTATAGGCTTAAAGCAAGCGGATTTCAAGAGTGAAGCAAGCTTACCGCAT 180
Db 790 AlaValSerValAlaLySphIleAlaCyAspPheSerGlyValGluGlnAlaLeuAlaSer 809
QY 181 CTCAAAATTTTTCAAAGAGCAATTTGCGCCCAACAGCTCAAAATAAGATGTCAT 240
Db 810 LeuLyAsnPheSerLySphGlnLeuAlaGlnAlaGlnIlySphAsnIlySerPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCATCGCTTAAGAAATGCTGAATGAAACCTTATGCT 300
Db 830 Val---GlyLySerGluLeuIleTyrGlnSerValIySphAsnGlyAlaAsnGlyThrLeuVal 848
QY 301 GGTAAAGGTATATCTCAAGCAAGCAACTCTTCTTAAATTTTGGACATCAAG 360
Db 849 GlyAsnGlyLeuSerGlyIleGlnAlaThrAlaLeuAlaLyAsnPhenSerPheIleAsn 868
QY 421 -----GAACCATTTATGTCTAAAGTTATTAATTAAGAAAGGCAAGGCAAGCTAGCCTT 474
Db 889 GlyGlyGluProIleTyrAlaGlnValAsnLySphThrGlyGlnValAlaSerPro 908
QY 475 GAAGAACCATTTACGCTCAAGTCTTAAAGGTAATGCAAAATTTGACGACTCAAT 534
Db 909 GluGluProIleTyrAlaGlnValAlaLySphValThrLySphIleAsnGlnLeuAsn 928
QY 535 CAATTAAGCA---AGTGGTTGGCTGTGAGGCAAGCA----- 570
Db 929 GlnAlaAlaThrSerGlyPheGlyGlyValGlyGlnAlaGlyPheProLeuLySphArgHis 948
QY 570 ----- 570
Db 949 AspLySphValGluAspLeuSerLySphValGlyArgSerValSerProGluProIleTyrAla 968
QY 571 -----GCGGAGCTTCCTCTTGAAGAGCAATGAATGATGATGCTC 612
Db 969 ThrIleAspAspLeuGlyGlySerPheProLeuLySphArgHisAspLySphValAspAspLeu 988
QY 613 AGTAAGGTAGGGCTTTCAAGAAATCAAGAAATGGCTCAAGAAATTTGACATCTCAAA 672
Db 989 SerLySphValGlyLeuSerArgAsnGlnGluLeuThrGlnIlySphIleAsnLeuSerGln 1008

APPLICATION NUMBER: US/08/316,397
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: SPRATT, GRENDOLYN D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.030
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1181 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-259-437-4

Alignment Scores:
 Pred. No.: 9e-85 Length: 1181
 Score: 902.50 Matches: 195
 Percent Similarity: 76.1% Conservative: 6
 Best Local Similarity: 73.9% Mismatches: 26
 Query Match: 77.1% Indels: 37
 DB: 2 Gaps: 4

US-09-360-685C-26 (1-685) x US-09-259-437-4 (1-1181)

```

QY 1 AAAAATGGCAAAATTAAGATTTTCAGCAAGTAAACGCAAAAGCGACTTGAAT 60
DB 750 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerApgGlnLys 769
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGGTAAACGATTAATTCATCAATCA 120
DB 770 SerLeuYsAenPheValIleIleAenGlnLysLysValThrApgLysLysAenGln 789
QY 121 GCGGTATCAGTGGCTTAAGCAACGGGTATTCAGTAAAGGTAAAGCAAGCGTAA 180
DB 790 AlaValSerValAlaLysIleAlaCysAenPheSerGlyValGlnGlnAlaLeuAla 809
QY 181 CTCAAAATTTCTCAAAAGCAATTCGCGCAACCACTCAAAAATTAAGATTCATCA 240
DB 810 LeuLysAenPheSerLysGlnLysLysValThrGlnAlaLysLysAenGlnLys 829
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGAAATGTTGAATGAAACCTG 300
DB 830 Val---GlyLysSerGlnLysLysValThrGlnSerValLysAenGlyLysAenGln 848
QY 301 GGTAAATGGTTATCTCAACAGAAAGCAACCTCTTCTAAAAATTTTCGCACTCA 360
DB 849 GlyAenGlyLysSerGlyIleGlnAlaThrAlaLeuAlaLysAenPheSerLys 868
QY 361 AAAGATTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAAC 420
DB 869 LysGlnLeuAenGlnLysPheLysAenPheAenAenAenAenAenAenAenAen 888
QY 421 -----GAACCATTTAAGTTAAGTTAATTAAGAAAGCAAGCGCAAGCTT 474
DB 889 GlyGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 908
QY 475 GAAGAACCACTTACGCTCAAGTTCCTTAAGAAAGGTAAATGCAAAATTAAGCA 534
DB 909 GlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 928
QY 535 CAATTAAGCA---AGTCGTTGGGTGGTGGGCAAGCA----- 570
DB 929 GlnAlaAlaThrSerGlyPheGlyGlyValGlyGlnAlaGlyPheProLeuLysA 948
QY 570 ----- 570
DB 949 AspLysValGlnAenPheSerLysValGlyAenSerValSerProGlnProLys 968
QY 571 -----GCGGGCTTCCTTTGAAAAGGCAATTAAGTTGATGATCTC 612

```

```

DB 969 ThrLeuAenPheLysGlyLysSerPheProLeuLysAenGlnAlaAspLysValA 988
QY 613 AAGATGAGGCTTTCAGAAATTCAGAAATTCAGAAATTCAGAAATTCAGAAATTC 672
DB 989 SerLysValGlyLysSerAenGlnLysLysValThrGlnAlaLysSerApgGln 1008
QY 673 GCGGTATCAGAA 684
DB 1009 AlaValSerGln 1012

RESULT 14
PCT-US93-09782-4
Sequence 4, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GRENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-4

Alignment Scores:
Pred. No.: 9e-85 Length: 1181
Score: 902.50 Matches: 195
Percent Similarity: 76.1% Conservative: 6
Best Local Similarity: 73.9% Mismatches: 26
Query Match: 77.1% Indels: 37
DB: 4 Gaps: 4

US-09-360-685C-26 (1-685) x PCT-US93-09782-4 (1-1181)
QY 1 AAAAATGGCAAAATTAAGATTTTCAGCAAGTAAACGCAAAAGCGACTTGAAT 60
DB 750 LysAenGlyLysAenLysAenPheSerLysValThrGlnAlaLysSerApgGlnLys 769
QY 61 TCCGTTAAAGATGTCATCATCAATCAAAAGGTAAACGATTAATTCATCAATCA 120
DB 770 SerLeuYsAenPheValIleIleAenGlnLysLysValThrApgLysLysAenGln 789
QY 121 GCGGTATCAGTGGCTTAAGCAACGGGTATTCAGTAAAGGTAAAGCAAGCGTAA 180
DB 790 AlaValSerValAlaLysIleAlaCysAenPheSerGlyValGlnGlnAlaLeuAla 809

```



```

QY 181 CTCGAAATTTCTCAAGAGCAATTCGCCCAAGCTCAAAAAATGAACTCAAT 240
DB 810 LeuLysAsnPhseSerLysGluGlnLeuAlaGlnAlaGlnLysAsnGlnSerPheAsn 829
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGAAAGGTGTGAATGAACTTATC 300
DB 830 Val---GlySerSerGlnLysGlnSerValLysAsnGlnValLysGlnThrLeuVal 848
QY 301 GGTATGGGTTATCTCAAGCAAGGCAACCTCTTCAAAAATCTTGGGCAATCAAG 360
DB 849 GlyAsnGlyLeuSerGlnLysGlnAlaThrAlaLeuAlaLysAsnPhseSerPheLys 868
QY 361 AAGAGTTGATGCAAACTTGAAATTTCAATATCAATATCAATATGAACTCAAAAC 420
DB 869 LysGlnLeuAsnGlnLysPheLysAsnPhseAsnAsnAsnAsnGlnLysAsn 888
QY 421 -----GAACCCATTTATGCTAAAGTTATTAATAAAGAAAGGAGGCAAGCTT 474
DB 889 GlyGlyGlnProLysGlnAlaGlnValAsnLysLysThrGlnValAlaSerPro 908
QY 475 GAAGACCATTTAGCTCAAGTTCGTAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 534
DB 909 GlnGlnProLysGlnAlaGlnValAlaLysLysValThrLysLysLysLysLysLys 928
QY 535 CAAATAGCA---AGTGGTTGGGTGTGTGAGGCAAGCA----- 570
DB 929 GlnAlaAlaThrSerGlnPheGlyGlyValGlyGlnAlaGlyPheProLeuLysArgHis 948
QY 570 ----- 570
DB 949 AspLysValGlnAspLeuSerLysValGlyArgSerValSerProGlnProLysGlnAla 968
QY 571 -----GCGGCTTCCTCTTGAAGGCAATGATTAAGTGAATGATTC 612
DB 969 ThrLeuAspLeuGlnLysSerPheProLeuLysArgHisAspLysValAspLeu 988
QY 613 AGTAAGTGAAGGCTTCAAGAAATCAAGAAATGCTCAAGAAATGCAATCAATCAATCA 672
DB 989 SerLysValGlyLeuSerArgAsnGlnLysLeuThrGlnLysLysLeuAsnLeuSerGln 1008
QY 673 GCGGTATCAGAA 684
DB 1009 AlaValSerGln 1012

```

RESULT 15

US-08-477-451-4
Sequence 4, Application US/08477451
Patent No. 5928865

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,451

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0335.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-4

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5,746-74	798.50	84.1%	84.1%	68.2%	3177	191	0	5	31	21

US-09-360-685C-26 (1-685) x US-08-477-451-4 (1-3177)

```

QY 3 AAATGGCAAAATATAGATTTTCAGCAAGTAAAGCAAGCAAAAGCACTTGAATTC 62
DB 2217 LysTrpGlnLys---GlyPheGlnGlnLysAlaSerLysLysArgPro---LysPhe 2234
QY 63 GGTAAAGATGTATCATCAATCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 122
DB 2235 Arg---ArgCysAspHisGlnSerLysGlnLysGlnLysGlnLysGlnLysGlnLys 2250
QY 123 GGTATCATGCTTAAGCAAGGCTGATTTTCAGTAAAGTAAAGTAAAGTAAAGTAAAGT 182
DB 2251 GlyLysSerGly---SerAsnGly---PheGln---GlyArgAlaSerValSerArgSer 2267
QY 183 CAAATATTTCTCAAGAGCAATTCGCCCAAGCTCAAAAATGAAAGTCAATG 242
DB 2268 GlnLysPheLeuLysGlyAlaLysGlyProThrSerSerLysLys---LysSerGlnCys 2286
QY 243 TAGAAAAAATCTGAATATATCAATCCGTTAAGAAAGTGTGAATGAAACCTTATGCG 302
DB 2287 ---LysLysLys---AsnLysSerLysLys---GlnTrpCysGlnTrpAsnProSerArg 2303
QY 303 TAATGGGTTATCTCAAGCAAGGCAAGCAAGCTTCTTCAAAAATCTTGGGCAATCAAGAA 362
DB 2304 ---TrpValLysSerSerArgSerHisAsnSerPhe---LysLeuPheGlyHisGlnGln 2321
QY 363 AAGATTGAATGCAAACTTGAAATTTCAATATCAATATCAATATGAACTCAAAAACGA 422
DB 2322 ArgValGlnCysLysTrpLysPheGln-----GlnGlnTrpPheGlnLysArg 2338
QY 423 ACCATTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAAGCTTGAAGAAC 482
DB 2339 ThrHisLeuCys-----SerLysGlnSerArgHisAsnSer---Pro---ArgThr 2353
QY 483 CATTTAGCTCAAGTTGTTAAAGGTTAATGCAAAAATGCAAGCACTCAATCAATATG 542
DB 2354 HisLeuArgSerSerCys---LysGlyLysCysLysAsn---ProLysGlnSerAsnSer 2371
QY 543 AAGTGGTTGGGTGTGTGAGGCAAGGCAAGGCTTCCTTGAAGGCAATTAAGT 602
DB 2372 LysTrpPheGlyCysCysArgAlaSerSerGlyLeuProPheGlnLysVala----- 2388
QY 603 TGATATCTCAGTAAAGGCTTTCAGGAATCAAGAAATGCTCAAGAAATGCAAA 662
DB 2389 -----SerSerGlnGlyArgAlaPheLysGlnSerArgLysGlnSerGlnLysAsn---Gln 2405
QY 663 TCTCAATCAAGCGGTATCAGA 683
DB 2406 SerGlnSerSerGlyLysArg 2412

```

Search completed: March 6, 2006, 20:02:57
Job time : 56.5 secs

This Page Blank (uspto)

XX

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.rag

Page 2

PN WO9318150-A1.
 XX
 PD 16-SEP-1993.
 XX
 PF 02-MAR-1993; 93WO-EP000472.
 XX
 PR 02-MAR-1992; 92IT-F1000052.
 PR 25-JAN-1993; 93WO-EP000158.
 XX
 PA (ISTS) BIOCIENCE SCLAVO SPA.
 XX
 PI Covacci A, Bugnoli M, Telford J, Macchia G, Rappuoli R;
 XX
 XX WPI, 1993-303464/38.
 DR
 DR N-PSDB; AAQ48733.
 XX
 PT Recombinant *Helicobacter pylori* protein and corresp. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and preventing
 PT type B gastritis, gastric ulcers and gastric tumours.
 XX
 PS Claim 5; Fig 4; 83pp; English.
 XX

This sequence represents the cytotoxin-associated aminoundulant antigen, CAI, of *H. pylori*. The protein is a hydrophilic, surface-exposed protein, having a molecular weight of approximately 120-132 kD, pI of 128-130 kD, and an isoelectric point of 9.72. *H. pylori* isolates which do not produce the CAI antigen do not have the cal gene, and are also unable to cal gene and cytotoxin. The association between the presence of the necessary for the transcription, folding, export or function of the cal gene is cytotoxin. Alternatively, both the cal gene and the c gene are absent in noncytotoxic strains [1]. In the present study, the physical linkage between the absence of a typical leader sequence suggests the presence of an independent export system. Sequence analysis suggests the presence of an region upstream from the AUG codon identified sequences resembling either the -10 or -35 regions, however, a region with a good consensus to the *E. coli* or *H. pylori* promoter sequence was not found. The CAI antigen is very hydrophilic and does not show obvious leader peptide or transmembrane sequences. The most hydrophilic region is from amino acids 600-900, where a number of unusual features are observed. There is a repetition of the sequences EFENKNDK and PLYN, and the presence of a stretch of six consecutive asparagines. This proline, and others derived from *H. pylori*, AAR12001, may be used to treat a heat shock protein (see also AAR1198 and infection). *H. pylori* is the causative agent of type B gastritis, peptic ulcers and gastric tumours. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)

Alignment Scores:	
Pred. No.:	7,156-106
Score:	1127.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	96.2%
DB:	2
Length:	1147
Matches:	228
Conservative:	0
Mismatches:	0
Indels:	0

US-09-360-685C-26 (1-685) x AAR41199 (1-1147)

QY	1	AAAAATGGCAAAATAATGAGCTTTTCAGCAGAGGTACCGCAAGCAAAAAGCAGCTTGA	60
Db	750	lysandilyysashulysapsheseriyetlaThnglnalysersapshleuglun	765
QY	61	TCCGTAAAGATGCGATCATCATTAAGAGGTACCGTAAAGCTGATTAATCTCAATCA	120
Db	770	serialysapvalleilelensnlnysaltthnashpyvalysapshleuansn	785
QY	121	CGCGTATCGTGGCGTAAAGCAGCGGATTTCTAGTACGGTAGACGACCGTTAGCGCAT	180
QY	790	AlataSerValAlaIysalathrcIysapsheserayvalnglnalaleuallap	805

QY	181	CTCAAAATTTCTCAAGAGCAATTCGGCCCAACGCTCAAAAAATGAAGTCAAT	240
Db	610	LeuIysAsnPhSerIysGlnIleuAlaGlnGlnIalGlnIlyIaAsnGlnSerLeuAsn	829
QY	241	CGTGAAGAAAAATTCGAAATATATCAATCCGTTAAGAAATGCTGAATGGAACCCAGTC	300
Db	830	AlaIleIysIysSerGlnIleIyGlnSerValIysAsnDlYValaGlnIyThrIleuVal	849
QY	301	GCTAATGGGTTATCTCAAGCAAGAACCCACAACTCTTCTTAAAACTTTTGGACATCAAG	360
Db	850	GlyAsnGlyIysSerGlnIalGlnIalThrThrIleuSerIyAsnPhSerAspIleIyAsn	869
QY	361	AAAGAGTTGAATGCAAAAATTCGAAATTTCAATACAAATACATTAATAGCTCAAAAAAC	420
Db	870	LysGlnIleuAsnAlaIysIeuGlyIaAsnPhSerAsnAsnAsnAsnGlnIyLeuIysAsn	889
QY	421	GAACCCATTTATGCTAAAGTAATATAAAAAAGAAAGGGCAAGAGCTAGCTTGAGAA	480
Db	890	GlnProIleIyAlaIyAlaIyAsnIyIysIyAlaGlyGlnIalAlaSerIeuGlnI	909
QY	481	CCCATTTACGCTCAAGTTGCTAAAAAAAGGTAAATGCAAAAATTTGACGATCAATCAATA	540
Db	910	ProIleIyAlaGlnIValAlaIyIysValaIleuAlaIyIleAspIyIleuAsnGlnIle	929
QY	541	GCAAGTGGTTGGGCTGTGAAGGCAAGCAAGCGGCTCCCTTTGAAAAAGCATGATAAA	600
Db	930	AlaSerGlyIeuDlYValIValIyGlnIalIalGlyPheProIleuIyAsnIleAspIyAsn	949
QY	601	GTTGATGATTCAGTAAGATGAAGGGCTTCAAGGAATCAAGAAATTCGTCAGAAATTCAG	660
Db	950	ValaIaAspIleuSerIyValIyGlyIeuSerIyAsnGlnIleuAlaGlnIyIleAsp	969
QY	661	AATTCATATTAAGCGGTATCAGA 684	
Db	970	AsnLeuAsnGlnAlaIValSerGln 977	
RESULT 2			
AAAB0600	ID	AAAB0600 standard; protein; 1338 AA.	
XX	XX	AAAB0600;	
AC	XX		
DT	XX	17-OCT-2003 (revised)	
DT	XX	15-FEB-1999 (first entry)	
XX	XX		
DE	XX	Helicobacter pylori Caga-Vibrio cholerae A2+B subunit fusion.	
XX	XX		
KW	XX	Vaccine; antigen; Caga; antigen; toxin; diagnosis; gastritis; ulcer;	
XX	XX	stomach cancer.	
OS	XX	Helicobacter pylori.	
OS	XX	Vibrio cholerae.	
OS	XX	Chimeric.	
XX	XX		
FN	XX	Key	Location/Qualifiers
FT	XX	Protein	1..1147
FT	XX		/label= CagA
FT	XX	Protein	1148..1338
FT	XX		/label= A2+B_subunits
XX	XX	MO9844130-A1.	
XX	XX	08-OCT-1998.	
PD	XX		
XX	XX	31-MAR-1998;	98NC-KR000073.
XX	XX		
PR	XX	31-MAR-1997;	97KR-00011950.
PR	XX	31-MAR-1997;	97KR-00011951.
XX	XX		
PA	XX	(DAEW-) DAEWOONG PHARM CO LTD.	
XX	XX		
XX	XX	Kim B, Shin S, Yu Y, Park M, Choi D, Jung H;	

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.rag

Page 4

XX	Sequence 1186 AA;
50	
Alignment Scores:	
Pred. No.:	9,086-89
Score:	959.50
Percent Similarity:	78.4%
Best Local Similarity:	77.7%
Query Match:	81.9%
DB:	8
Length: 1186	
Matches:	205
Conservative:	2
Mismatches:	20
Indels:	37
Gaps:	3

US-09-360-685C-26 (1-685) x ADU05584 (1-1186,

QY	1	AAAAATGGCAAAAATTAAAGTTTTCAGCAAGCTTAACGCAAGCAAAAACGCACTTGAAAT	60
Db	755	LYaBaNGLyBaenlyBaPhPheSerLyValThnGlnAlaLysSerAspLeuGlnaBn	77
QY	61	TCCGTTAAAGATGATCATCATCAAAAGCTAACGCTAAAGTTGATATCTCAATCA	120
Db	775	SeLyValyBaPValIlelleaBnInLyValThnZnAspLyValAspAsnLeuAaGln	794
QY	121	GGCGATTACAGTGGCTAAACCAACGGGTGATTTCTAGAGGTATAGCAAGCCTTAGCGCAT	180
Db	795	AlaValSerValAlaLyValAlaWeCdyAspPheSerZnLyValGlnGlnValLeuAlaBaP	814
QY	181	CTCAAAAATTTCTCAAAAGAGCAATTGGCCCAACAGCTCAAAAATAAGAAAGTCCAT	240
Db	815	LeuLyBaBnPhPheSerLySeGlnLeuAlaGlnGlnAlaGlnLyBaBnGLyBaPheAaBn	833
QY	241	GCTAAGAAAAATCTGAATATATCATCCGTTAAGAATGGTGTGAATGGCAACCTTAGTC	300
Db	835	ThnGlyLyBaBnSeGlnLeuTyGlnSerValLyBaBnSerZnAlaLysThnLeuVal	854
QY	301	GGTAAATGGGTTATCTCAAGCAAGCAACCAATCTCTTCTTAAATACTTTGGGATCAAG	360
Db	855	GlyAaBnGLyLeuSeGlnYlIeGlnAlaThnAlaLeuAlaLyBaBnPhPheSerAaPheLyS	874
QY	361	AAGAGTTGAATCAAAAATCTGGAATTTCAATAAATCAATCAATATGATCAAAAC	420
Db	875	LySGLyLeuAaBnGLyBaPheLyBaBnPhPhe--AaBnBaBnAaBnGLyLeuAaBn	893
QY	421	-----GAAACCAATTTAGCTAAAGTTAATAAAAGAAACAGGSCAAGCAAGCTTAGCCTT	474
Db	894	SeThnGlnProlIeTyAlaLyValAaBnLyValySerThnGlyGlnAlaAlaSerPro	913
QY	475	GAAAGAACCCATTAGCTCAAGTTGCTAAAAAGGTAATGCAAAAATTGACACGATCCAT	534
Db	914	GlnGluProlIeTyThnGlnValAlaLyValAaBnAlaAlaLySilleBaPheGlnaBn	933
QY	535	CAAAATACCAAGTGGTTGGGTGTGTAAGGGCAACAGCCGGCTCCCTTGAAGAAAGCAT	594
Db	934	GlnIleAlaSeGlnLyLeuGlyGlyValAlaGlyAlaAlaGlyPheProlIeTyAaBnS	953
QY	595	GATTAAGTTGATGATCTTCAGTAAGTAGAGGCTT-----	627
Db	954	AaPlyValAaBnAspLeuSerLyValAlaGlyLeuSerAlaSerProGlnProlIeTyAla	973
QY	627	-----	627
Db	974	ThnIleAspAspLeuGlyGlyProPheProlIeTyAlaGlnHisAaBnLyValAaBnAspLeu	993
QY	628	-----TCAAGAAATCAAAATTTGGCTCGAAATAATGACCAATCTTCAATCA	672
Db	994	SeLyValAlaGlyAaGSerZnLyBaBnGlnLeuAaGlnLySilleBaPnLeuAaGln	1013
QY	673	GGCGATTACGAA	684
Db	1014	AlaValSerGln	1017
RESULT 4			
ABU51614			
ID	ABU51614	standard; protein; 422 AA.	

AC ABB51614;
XX
XX 07-MAY-2003 (first entry)
DT
DT
XX Helicobacter pylori selected interacting domain (SID) protein #958.
DE
DE
XX Protein-protein interaction; ulcer; selected interacting domain; SID
XX
XX Helicobacter pylori.
XX

PN WO200266501-A2
XX

PD 29-AUG-2002

28-DEC-2001: 2001WO-FB01E438

2X
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

02-04M-2007; T007-4TH-20
XX XX

PA (HYBR-) HYBRIGENICS.

(INSE) INSI PASTEUR.

XX

PI Legrain P, Rain J, Colland F, De Reuse H, Labiane A:

WPT: 2002-674910/72

DR N-PSDB; ABX66359.

New complexes of pro-

PT useful for identifying modulating compounds for treating or preventing

ulcers in mammals.

PS Claim 6; Page 318; 642pp; English

The invention is described in the following figures:

CC Helicobacter pylori selected from 421 complexes given in the

specificity. The complex of protein-protein interactions are useful for screening for agents that

CC Modulating compounds which binds to a targeted bacterial protein.

used for treating or preventing ulcers in a human or animal. This is the amino acid sequence of a protein.

CC via protein-protein interactions. Note: Where the parent number is

at the top of the pages in the specification has obscured areas of protein sequence that is

CC represent an illegible residue

XXXXXXXXXXXX

Alignment Scores:

Score:	2:205-03	Length:	422
	905.50	Matches:	105

Percent Similarity: 77.6%

Best Local Similarity: 75.0%

Conservative: 2

```

-----
Query Match: 77.38
Mismatches: 20
Indels: 27

```

DE: 5
Gaps: 3

US-09-360-685C-26 (1-685) x ABU51614 (1-422)

22 **=====**

CGATGCTTGAAGATGTGATCATCAACAAG 90

Db 1 ValThrGlnAlaLysSerAspLeuGluAsnSerValLysAspValIleIleAsnGlnLys 20

91 GTACCGATAAAGTTGATAATTCAATCAGCCGCCGTCTTCACC

150 100 50 0

22 val...mbpLybvalabpabnLeuabnglnAlaValSerValAlaLysAlaMetGlyAsp 40

151 TTCACTAGGGTAGAGCAAGCGTTAGCCGATCTCAAAATTTCTCAAAGGAGCAATTGCC 210

[illegible]

60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074

211 CAAACAGCTCAAAAATGAAGCTCTCAATGCTAGAAATAATCTGAATATATCAATCC 270

61 G I N G I N A I a G I N L y S a S n G I u a S p p h e a S n T h r G I L y S a S n S e r G I u L e u T y r G I n S e r 80

271 GTTAAAGGATGGTGCTGAATCCCAACCGCACCACCGCAGGCAGG

Db 81 ValIySaSnSeRValIaSnIyStHrIeValGIyAsnGIySeuSeRGIyIIeGIuIaIaThr 1000

QY 331 ACTCTTTCTAAAAACTTTTTCGGACATCAAGAAAAAGCTTGAATGCMAAAACTTGAAATTTC 390

Db 101 AlAlaValAlaIySaSnPhSeSeRAspIIeIySlyGIyIleuSnngIuIySpHeIySaSnPhe 1200

QY 391 AATAACAATTAACAATAATGACACTCAAAAAC-----GAACCCATTATAGCTAAAGTTAAT 444

Db 121 ---AsaAsnAsnAsnAsnngIyIeuIySaSnSeRThrGIuProIleIyAlaIySaIAsn 139

QY 445 AAAAAAGAAACGAGGCGAAGCAGCTAGCTTGAAGAAACCCATTACGCTCAAGTGTCTAA 504

Db 140 IySlySlyStHrGIyGIuIuValIaIaSeRProGIuGIuProIleIyThrGIuIuValaIyS 159

QY 505 AAGGTAAATGCMAAAATTGACCGACTCAATCAAAATGCAAGTGTTGGTGTGTAGGG 564

Db 160 IySValaSnAlaIySIIeaSpArGIleuAnngIuIleAlaSeRGIyIeuGIyIyValGIy 179

QY 565 CAAGCAGCGGCGCTCCCTTTGAAAAGCGATGATAAAGTTGATATCTCAGTAAAGTAGGG 624

Db 180 GIuAlaIaIaGIyPheProIeuIySaRGIaSaSpIySaIaSaSpSeuSeRGIySValGIy 199

QY 625 CTT----- 627

Db 200 IeuSeRAlaSeRProGIuProIleIyAlaThrIIeaSaSpIeuGIyGIyProPhePro 219

QY 628 -----TCAAGAAATCAAGA 642

Db 220 IeuIySaRGIaSaSpIySaIaSaSpIeuSeRGIySaIaIyYArGeSeRArGaAnngIu 239

QY 643 TTGGCTCAGAAAAATTGACAATCTCAATCAAGGGGTATTCGAA 684

Db 240 IeuAlaGIuIySIIeaSpAsnIeuAnngIuAlaIaIaSeRGIu 253

Tue Mar 7 12:58:24 2006

us-09-360-685c-26.rag

Page 6

AA072593	standard; protein; 1181 AA.
AA072593	
25-MAR-2003	(revised)
29-SEP-1995	(first entry)
H. pylori tagA antigen.	
TagA; antigen; ulcer; diagnosis; vaccine.	
Helicobacter pylori.	
US5403924-A.	
04-APR-1995.	
26-APR-1993;	93US-00053614.
13-OCT-1992;	92US-00959940.
(UUYA-) UNIV VANDERBILT.	
Blaser MJ, Tummuru MKR, Cover TL;	
WPI; 1995-146855/19.	
N-PSDB; AA086728.	
New nucleic acid encoding tag A antigen of Helicobacter pylori - used to detect preinvasion to peptic ulceration and to produce protein for use in vaccine; diagnosis etc.	
Disclosure; Col 37-46; 30pp; English.	
The full-length sequence of the tagA gene of H. pylori 84-183 (ATCC 53726), was obtained from overlapping clones isolated from genomic libraries. The gene encoded a 1181-amino acid TagA antigen protein. (updated on 25-MAR-2003 to correct PF field.)	
Sequence 1181 AA;	

Alignment Scores:	
Pred. No.:	5,989-83
Score:	900.150
Percent Similarity:	76.150
Best Local Similarity:	73.98
Query Match:	77.18
DB:	
US-09-360-685C-26	
(1-685) x AAR72593	
(1-1181)	
Length:	1181
Matches:	195
Conservative:	6
Mismatches:	26
Indels:	37
Gaps:	4

Qy	AAAAATGGCAAAAATTAAGGATTTTTCAGCAAGTAAACGCAACAAAAGGACCTTGAAAT	60
Db	lyeaenglyyaleuylayepheseSerValThrglnAlaySeAepGnglnuhen	765
Qy	61 TCCGTTAAAGATGTATCATCATCAAAAAGTAAACGATAAAGTTGATATCTCAATCA	1295
Db	770 SerIlyeasapValIleIleahenglnlybIethmrpyleValAspGlnleuAengln	785
Qy	121 GGGGATCAATGGCTCTAAAGCAACGGTGAATTTTCAGTGGGTAGACCAAGCTTAGCCAT	1805
Db	790 AlaValSerValAlaIyAlIleAlaIyAclyasepheseerIyValGlnGlnAlaIleuAlaAsp	805
Qy	181 CTCAAAATTTTCAAAGAGCAATTTGGCCCAAGCAAGCTCAAAAATTAAGAAAGTCTCAAT	240
Db	810 leuylasrnpheSerlyeGlnGlnleuAlaglnlnlnGlnlyyaseglnsrphehen	825
Qy	241 GCTAGAAAAAATTCGAATATATTCATTCCTTTAAGATAGTGGTGAATGAACCTTAGTC	300
Db	830 Val---GlyValSeeglnIleIyrglnSerValIyAenglyValaenglnymrIleuVal	848
Qy	301 GGTATAGGTTATCTCAAGCAAGACCAACACTTTCTTAAAAACCTTTGGACATCAAG	360

Db 849 Gly¹Met²Gly³Leu⁴Ser⁵Gly⁶Leu⁷Ala⁸Val⁹Ala¹⁰Thr¹¹Ala¹²Leu¹³Val¹⁴Ser¹⁵Pro¹⁶Ile¹⁷Leu¹⁸ 866

QY 361 AAAAGCTTGAATGCAAAACCTTGAAATTCATTAACAATTAACAATTAATGCACTCAAAAC 420

Db 869 Lys¹Glu²Leu³Met⁴Gln⁵Ile⁶Pro⁷Leu⁸Phe⁹Leu¹⁰Val¹¹Asn¹²Phe¹³Asn¹⁴Met¹⁵Asn¹⁶His¹⁷Leu¹⁸Val¹⁹Asn 886

QY 421 -----GAACCCATTATAGCTTAAGCTTAATTAATAAAAGAAACGAGGCCAAGCAAGCTACCTT 474

Db 889 Gl¹Gly²Ile³Pro⁴Ile⁵Thr⁶Ile⁷Val⁸Gln⁹Val¹⁰Leu¹¹Asn¹²Val¹³Ser¹⁴Thr¹⁵Gln¹⁶Val¹⁷Ala¹⁸Ser¹⁹Pro 908

QY 475 GAAACAACCACTTACGGCTCAAGTTCGTTAAAAAGCTAAATGCAAAAATTAATGACCAAGCTCAT 534

Db 909 Gl¹Ile²Pro³Ile⁴Thr⁵Ile⁶Gln⁷Val⁸Ala⁹Val¹⁰Ala¹¹Val¹²Ser¹³Val¹⁴Thr¹⁵Val¹⁶Ser¹⁷Ile¹⁸Asp¹⁹Gln²⁰Leu²¹Asn 928

QY 535 CAATATACCA---AGTCGTTGGCGGTGTTGACGGCAAGCA--- 570

Db 929 Gl¹Ala²Ile³Ala⁴Thr⁵Ser⁶Gly⁷Phe⁸Gly⁹Ile¹⁰Val¹¹Gln¹²Ile¹³Gly¹⁴Phe¹⁵Pro¹⁶Leu¹⁷Val¹⁸Arg¹⁹His 948

QY 570 ----- 570

Db 949 Asp¹Leu²Val³Glu⁴Asp⁵Leu⁶Ser⁷Ile⁸Val⁹Gly¹⁰Arg¹¹Ser¹²Val¹³Ser¹⁴Pro¹⁵Glu¹⁶Pro¹⁷Ile¹⁸Thr¹⁹Ala 968

QY 571 -----GGCGGCTTCCTCTTTGAAAAGGCACTGTAATGATGATGATCTC 612

Db 969 Thr¹Ile²Asp³Asp⁴Leu⁵Gln⁶Gly⁷Ser⁸Phe⁹Pro¹⁰Leu¹¹Val¹²Arg¹³His¹⁴Asp¹⁵Leu¹⁶Val¹⁷Asp¹⁸Asp¹⁹Leu 988

QY 613 AGTAAGGTAGGGCTTTCAAGAGAAATTCAGAAATTTGGCTCAAGAAATTTACATCATCTCAATCA 672

Db 989 Ser¹Leu²Val³Gly⁴Leu⁵Ser⁶Arg⁷Asn⁸Gln⁹Ile¹⁰Leu¹¹Thr¹²Gln¹³Ile¹⁴Val¹⁵Ile¹⁶Asp¹⁷Asn¹⁸Leu¹⁹Ser²⁰Gln 1006

QY 673 GCGGTCATCAAG 684

Db 1009 Ala¹Val²Ser³Gln 1012

```

RESULT 7
AAR91307
ID      AAR91307  standard; protein; 1181 AA.
XX

```

AC	AA91307;	
XX	09-SEP-1996	(first entry)
DT		
XX		
XX	Helicobacter pylori TagA 120-128 kD antigen.	
XX		
DE	Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;	
XX	diagnosis; predisposition; antibody; vaccine; infection.	
KW	Helicobacter pylori.	
OS		

PN WO9610639-A2.
XX

PD 11-APR-1996.
XX

29-SEP-1995; 95WO-US012669.

30-SEP-1994; 94US-00316397.

(UYVA-) UNIV VANDERBILT.
(ORAY-) ORAYNY TWO
PA
PA

Cover Title: PJ

cover ID, Blaser MJ, Kleanthous H, Tummuru MKR, WPI : 1995-2003/01

N-PSDB; AAT14051.

Helicobacter pylori

XX treatment and prevention of peptic ulceration and gastric carcinoma.

Claim 4; Page 103-107; 118pp; English.

AA91307 is a 120-128 kD antigen of *Helicobacter pylori* (Hp) the anti-

CC is designated TagA and its, or its fragments, can be used to determine a
 CC predilection to peptic ulceration or gastric carcinoma, conditions
 CC caused by or linked to HP infection. TagA and its fragments may also be
 CC used for antibody prodn. for use in detection of TagA in patients
 CC suspected of HP infection. TagA antibodies or other ligands may also be
 CC used to treat peptic ulceration or gastric carcinoma caused by HP
 CC infection. TagA or a non-functional TagA mutant may be used in vaccines
 CC for preventing and treating HP infection

XX Sequence 1181 AA:

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	5,98e-83	1181	195	6	26	37
Percent Similarity:	902.50					
Best Local Similarity:	76.1%					
Query Match:	73.9%					
DB:	77.1%					

US-09-360-685C-26 (1-685) x AAR91307 (1-1181)

QY 1 AAAAATGGCAAAATAGATTTTCAGCAAGCTAACGCAAAAAGCGACTTGAAT 60
 DB 750 LysAsnGlyLysAsnLysAspPheSerLysValThrGlnAlaLysSerAspGlnLysVal 769
 QY 61 TCCGTTAAAGATGATCATCAATCAAAAGCGATTAAGTGAATATTCATCA 120
 DB 770 SerIleLysAspValIleIleAsnGlnLysIleThrAspLysValAspGlnLysVal 789
 QY 121 GCGGTATCAAGTGGCTAAAGCAAGCGGTGATTTTCAGTAGAGCAAGCGTGA 180
 DB 790 AlaValSerValAlaLysIleAlaCysAspPheSerGlyValGlnGlnAlaLeuAlaAsp 809
 QY 181 CTCGAAAATTTCTCAAGAGCAATGGCCCAACAAGCTCAAAAAATGAAGTCTCA 240
 DB 810 LeuLysAsnPheSerLysGlnGlnLysValGlnGlnAlaGlnLysAsnGlnSerPheAsn 829
 QY 241 GCTAGAAAAAATCTGAATATTCATCCGTTAAAGATGCTGAATGAAACCTGATC 300
 DB 830 Val--GlyLysSerGlnIleLysValLysValLysValLysValLysValLysVal 848
 QY 301 GGTATGGCTTATCTCAAGCAAGCAAGCAAGCTTCTTCAAAAATTTTCGACATCA 360
 DB 849 GlyAsnGlyLysSerGlyIleGlnLysIleThrAlaLeuAlaLysAsnPheSerPheLys 868
 QY 361 AAAAGTTGAATGCAAACTGGAAAATTTCAATTAACATTAACATTAATGACTCA 420
 DB 869 LysGlnLysAsnGlnLysPheLysAsnPheAsnAsnAsnAsnAsnGlnLysAsn 888
 QY 421 -----GACCATTATCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 474
 DB 889 GlyGlyLysIleProIleLysValGlnValLysValLysValLysValLysValLys 908
 QY 475 GAAGAACCCATTACGCTCAAGTGTCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 534
 DB 909 GlnGlnProIleLysIleLysValLysValLysValLysValLysValLysValLys 928
 QY 535 CAATATGCA--AGTGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 570
 DB 929 GlnAlaLysThrSerGlyPheGlyValGlnGlnAlaGlyPheProLysValLys 948
 QY 570 ----- 570
 DB 949 AspLysValGlnAspLysSerLysValGlyArgSerValSerProGlnProIleLysVal 968
 QY 571 -----GCGGGCTTCCTTTGAAAGGATGATTAAGTGAATGATC 612
 DB 969 ThrIleAspAspLysLysLysSerPheProLysValLysValLysValLysValLys 988
 QY 613 AGTAAGTGAAGGCTTCAAGCAATCAAGATTTGGCTCAGAAAATTTGACATCTCA 672
 DB 989 SerLysValGlyLysSerArgAsnGlnLysLysValLysValLysValLysValLysVal 1008

QY 673 GCGGTATCAGAA 684
 DB 1009 AlaValSerGln 1012

RESULT 8

ID ABU51367 standard; protein; 420 AA.

AC ABU51367;

DT 07-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (STD) protein #711.

KW Protein-protein interaction; ulcer; selected interacting domain; STD.

OS Helicobacter pylori.

PN MO200266501-A2.

PD 29-AUG-2002.

PF 28-DEC-2001; 2001WO-BP015428.

PR 02-JAN-2001; 2001US-0259302P.

PA (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.

PI Legrain P, Rain J, Collard F, De Reuse H, Labigne A;

DR WPI; 2002-674910/72.

DR N-PSDB; ABX6112.

PT New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.

PS Claim 6; Page 257; 642pp; English.

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (STD), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue

SO Sequence 420 AA;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1,89e-82	420	193	2	37	3
Percent Similarity:	896.50					
Best Local Similarity:	77.4%					
Query Match:	76.6%					
DB:	5					

US-09-360-685C-26 (1-685) x ABU51367 (1-420)

QY 37 CAAGCAAAAGCGACTTGAATTCGTTAAAGATGATCAATCAAAAGGTAC 96
 DB 1 GlnAlaLysSerAspLysGlnAsnSerValLysAspValIleIleAsnGlnLysValThr 20
 QY 97 GATTAAGTTGATTAATCTCAATCAAGCGGTATCAGTGTAAAGCAAGCGGTGATTCAGT 156
 DB 21 AspLysValAspAsnLysAsnGlnAlaValSerValAlaLysAlaMetGlyAspPheSer 40
 QY 157 AGGTTAGCAAGCGGTAGCCGATCTCAAAAATTTCTCAAGAGCAATGGCCCAACAA 216

Db 41 ArgValGluGlnValLeuAlaSerPheLeuValSerPheSerLeuGlnLeuAlaGlnGln 60
 Qy 217 GCTCAAAAATGAAATGCTCAATGCTAGAAAAATCTGAAATATCAATCCGTTAAG 276
 Db 61 AlaGlnPheSerLeuAlaSerPheSerLeuGlnLeuAlaGlnLeuAlaGlnLeuAla 80
 Qy 277 AATGCTGAAATGCAACCTTACGCTGTAATGCTTATCTCAAGCAAGCAACCACTT 336
 Db 81 AsnSerValAsnLeuPheThrLeuValGlnPheSerLeuGlnLeuAlaGlnLeuAla 100
 Qy 337 TCTAAAACTTTTCGACATCAAGAAAGTGAATGCAAACTTGAAATTCATTAAC 396
 Db 101 AlaPheSerLeuAlaSerPheSerLeuGlnLeuAlaGlnLeuAlaGlnLeuAla 119
 Qy 397 AATTAACATTAATGCACTCAAAAAC-----GAAACCACTTATGCTAAATTAAG 450
 Db 120 AsnAsnAsnAsnGlnLeuValSerPheSerLeuGlnLeuAlaGlnLeuAlaGln 139
 Qy 451 AAAGCAAGGCAAGCAAGCTTACGCTTGAAGAACCTTATGCTCAAGTGTAAAGAG 510
 Db 140 LysThrGlnGlnValAlaSerProGlnGlnProLeuThrGlnValAlaGlnLeuAla 159
 Qy 511 AATGCAAAATTTGACCGACTCAATCAATATGCAAGTGTGCTGTTAGGCAAGCA 570
 Db 160 AsnAlaLysLeuAlaSerPheSerLeuGlnLeuAlaGlnLeuAlaGlnLeuAla 179
 Qy 571 GCGGCTTCCTTGAAGAAAGCATTAAGTTGATGATGATGATGATGATGATGATGAT 627
 Db 180 AlaGlnPheProLeuLysArgHisAspLeuValAlaSerPheSerLeuGlnLeuAla 199
 Qy 627 ----- 627
 Db 200 AlaSerProGlnProLeuLysArgHisAspLeuValAlaSerPheSerLeuGlnLeuAla 219
 Qy 628 ----- 628
 Db 220 ArgHisAspLeuValAlaSerPheSerLeuValAlaGlnLeuAlaGlnLeuAla 239
 Qy 649 CAGAAATTTGACATGCTCAATCAAGCGGCTTCAAGAA 684
 Db 240 GlnLysLeuAlaSerPheSerLeuGlnLeuAlaGlnLeuAlaGlnLeuAla 251
 RESULT 9
 AEB91411
 ID AEB91411 standard; protein; 1167 AA.
 AC AEB91411;
 XX 20-OCT-2005 (first entry)
 DT 20-OCT-2005 (first entry)
 DE Microbial pathogen adhesin protein sequence. SEQ ID NO:121.
 XX
 KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;
 KW Bordetella pertussis infection; antibacterial; pneumonia;
 KW antiinflammatory; respiratory-gen.; gastric ulcer; antifungal;
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
 OS Helicobacter pylori.
 XX
 PN WO2005076010-A2.
 PD 18-AUG-2005.
 XX
 PR 07-FEB-2005; 2005WO-IN00037.
 PR 06-FEB-2004; 2004IN-DE000173.
 PR 20-FEB-2004; 2004US-0589227P.
 XX
 PA (COUL.) COUNCIL SCI & IND RES SOUTH AFRICA.
 XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramchandran S;
 XX

DR WPI, 2005-597835/61.
 XX
 PT Computational method for identifying adhesin and adhesin like molecules,
 PT comprises computing sequence-based attributes of protein sequences using
 PT neutral network software and training an artificial neural network.
 XX
 PS Claim 16; SEQ ID NO 121; 402pp; English.
 CC The present invention relates to a computational method (M1) for
 CC identifying adhesin and adhesin-like proteins, by computing the sequence-
 CC based attributes of protein sequences using five attribute modules of a
 CC neutral network software, training an artificial neural network (ANN) for
 CC each of the computed five attributes, and identifying adhesin and
 CC adhesin-like proteins having probability of being an adhesin (P_{ad}) and
 CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
 CC encoding adhesin and adhesin-like proteins, having 274 fully defined
 CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
 CC genes encoding adhesin and adhesin-like proteins, having 105 fully
 CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
 CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
 CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
 CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
 CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)
 CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
 CC proteins, of therapeutic potential, and identifying adhesin and adhesin-like
 CC proteins for further testing in development of new vaccine formulations
 CC to eliminate diseases caused by various pathogenic organisms. (M1) is
 CC useful for identifying putative adhesins that are important in drug
 CC discovery and preventing therapeutic for whooping cough, pneumonia,
 CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
 CC unique proteins. The present sequence is a microbial pathogen adhesin
 CC protein sequence.
 XX
 SQ Sequence 1167 AA;
 Alignment Scores:
 Pred. No.: 2 446-82 Length: 1167
 Score: 896.50 Matches: 189
 Percent Similarity: 74.88 Conservative: 7
 Best Local Similarity: 72.18 Mismatches: 15
 Query Match: 76.68 Indels: 51
 DB: 9 Gaps: 2
 US-09-360-685c-26 (1-685) x AEB91411 (1-1167)
 Qy 1 AAAAAATGCAAAATATGATTTTACAGGTAACGCAAGCAAAAGCACTTGAAT 60
 Db 754 LysAsnGlnLysAsnLysAspPheSerLeuValThrGlnAlaLysSerAspLeuGlnLeu 773
 Qy 61 TCGCTTAAGATGATGATCAATCAATCAAAAGGTAACGATTAAGTTGATTCATCA 120
 Db 774 SerLeuLysAspValAlaLeuSerGlnLysLeuThrAspLeuValAlaSerPheSerLeu 793
 Qy 121 GCGGTATCGGCTTAAGCAACGAGTATTCAGTAGGAGTGAAGCAAGCTTACCGAT 180
 Db 794 AlaValSerValAlaLysAlaThrGlnLysPheSerGlnValGlnGlnAlaLeuAlaSer 813
 Qy 181 CTCAAAAATTTCTCAAGAGCAATGCGCCCAACAGCTCAAAAAATTAAGTCTCAAT 240
 Db 814 LeuLysAsnPheSerLeuGlnGlnLeuAlaGlnGlnLysAsnGlnLysPheAsn 833
 Qy 241 GCTAGAAAAATCTCAAAATATATCAATCCGTTAAAGATGCTGAATGCAACCTTAC 300
 Db 834 ThrGlnLysAsnSerAlaLeuThrGlnSerValLysAsnGlnLysAlaSerGlnThrLeuVal 853
 Qy 301 GGTATGAGTTATCTCAAGAGAGCAACCACTTCTAAAACTTTTCGACATCAAG 360
 Db 854 GlyAsnGlnLysSerLeuValAlaThrThrLeuSerLysAsnPheSerAspLeuLys 873
 Qy 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGATGATCAAAAG 420

Db 874 LysGluLeuAsnAlaIysLeuGlyAsnPheAsnAsnAsnAsnGlyLeuGluAsn 893
 Qy 421 GAACCCATTATGCTAAAGTTAAATAAAGAGAGGAGAGAGCTACCTTGAAAGA 480
 Db 894 Ser-----ThrGlu 896
 Qy 481 CCCATTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACGACTCAATCAATA 540
 Db 897 ProIleTyrThrGlnValAlaIysValIleValAlaIysIleAspArgLeuAspGlnIle 916
 Qy 541 GCAGTGGTTGGGTGTTGAGGCAACGCGGCG----- 576
 Db 917 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuIysArgHisAspIys 936
 Qy 576 ----- 576
 Db 937 ValAspAspLeuSerIysValGlyLeuSerAlaAsnHisGluProIleTyrAlaThrIle 956
 Qy 577 -----TTCCCTTTGAAAAGGCGATGATTAAGTTGATGATCTCAGTAAG 618
 Db 957 AspAspLeuGlyGlyProPheProLeuIysArgHisAspIysValAspAspLeuSerIys 976
 Qy 619 GTAGGCTTTCAAGATCAAGATTCGCTCAGAAAATTGACATCTCAATCAAGCGGTA 678
 Db 977 ValGlyLeuSerArgGlnGlnIysLeuThrGlnIysIleAspAsnLeuAsnGlnAlaVal 996
 Qy 679 TCAGAA 684
 Db 997 SerGlu 998
 RESULT 10
 AAM20174
 ID AAM20174 standard; protein; 1178 AA.
 XX AAM20174;
 AC 08-JUL-1997 (first entry)
 DT
 XX
 DE H. pylori cytoplasmic protein, 16459375.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 XX
 PN W09640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US009122.
 XX
 PR 07-JUN-1995; 95US-00487032.
 PR 01-APR-1996; 96US-00630405.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Smith D, Berglindh OT, Mellgaard BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67410.
 XX
 PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 useful for vaccines to treat or prevent H. pylori infection, and to
 detect Helicobacter.
 PT
 XX
 PS Claim 61; Page 390-392; 1481pp; English.
 XX
 CC The present sequence is a Helicobacter pylori cytoplasmic protein. The
 CC protein may be used in a vaccine to prevent or treat H. pylori infection
 CC or to identify H. pylori polypeptide binding compounds, useful as
 CC potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping

CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts
 XX
 SQ Sequence 1178 AA;
 Alignment Scores:
 Pred. No.: 2,45e-82 Length: 1178
 Score: 896.50 Matches: 189
 Percent Similarity: 74.8% Conservative: 7
 Best Local Similarity: 72.1% Mismatches: 15
 Query Match: 76.6% Indels: 51
 DB: 2 Gaps: 2
 US-09-360-685C-26 (1-685) x AAM20174 (1-1178)
 Qy 1 AAAAAAGCAAAATAGGATTTCAGCAAGTACGCAAGCAAAAGCACTTGAATA 60
 Db 765 LysAsnGlyLysAsnIysAspPheSerIysValIleGlnAlaIysSerAspLeuGluAsn 784
 Qy 61 TCCGTTAAAGATGTGATCATCATCAATCAAAAGTACGATTAAGTTGATATCTCAATCA 120
 Db 785 SerIleLysAspValIleIleAsnGlnIysIleThrAspIysValAspAsnLeuAsnGln 804
 Qy 121 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTTCAATGAGGTATGACCAAGCTTACCGAT 180
 Db 805 AlaValSerValAlaIleValAlaThrIleAspPheSerGlyValGlnGlnAlaIleValAsp 824
 Qy 181 CTCAAAAATTCTCAAGAGCAATTGGCCCAACAGCTCAAAAAATGAAGTTCAT 240
 Db 825 LeuLysAsnPheSerIysGlnIleValIleGlnAlaGlnIleLysAsnGluAspPheAsn 844
 Qy 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATGGTGAATGAAACCTAGTC 300
 Db 845 ThrGlyLysAsnSerAlaLeuTyrGlnSerValIysAsnGlyValAsnGlyThrIleVal 864
 Qy 301 GGTATGGGTATTCAGACGACGACCAACTCTTTCTAAAACTTTGCGACATCAAG 360
 Db 865 GlyAsnGlyLeuSerIysAlaGlnAlaThrThrIleSerIysAsnPheSerAspIleLys 884
 Qy 361 AAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAACATTAAC 420
 Db 885 LysGluLeuAsnAlaIysLeuGlyAsnPheAsnAsnAsnAsnAsnGlyLeuGluAsn 904
 Qy 421 GAACCCATTATGCTAAAGTTAAATAAAGAGAGGCAAGGAGAGCTAGCTTGAAGA 480
 Db 905 Ser-----ThrGlu 907
 Qy 481 CCCATTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACGACTCAATCAATA 540
 Db 908 ProIleTyrThrGlnValAlaIleValIysValIleValIysIleAspArgLeuAspGlnIle 927
 Qy 541 GCAGTGGTTGGGTGTTGAGGCAACGCGGCG----- 576
 Db 928 AlaSerGlyLeuGlyAspValGlyGlnAlaAlaSerPheLeuLeuIysArgHisAspIys 947
 Qy 576 ----- 576
 Db 948 ValAspAspLeuSerIysValGlyLeuSerAlaAsnHisGluProIleTyrAlaThrIle 967
 Qy 577 -----TTCCCTTTGAAAAGGCGATGATTAAGTTGATGATCTCAGTAAG 618
 Db 968 AspAspLeuGlyGlyProPheProLeuIysArgHisAspIysValAspAspLeuSerIys 987
 Qy 619 GTAGGCTTTCAAGATCAAGATTCGCTCAGAAAATTGACATCTCAATCAAGCGGTA 678
 Db 988 ValGlyLeuSerArgGlnGlnIysLeuThrGlnIysIleAspAsnLeuAsnGlnAlaVal 1007

QY 679 TCAGAA 684
DB 1008 SerGlu 1009

RESULT 11

AAW20884
ID AAW20884 standard; protein, 1183 AA.

AC AAW20884;

DT 18-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 13ee12016orf74.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
diagnosis.

XX Helicobacter pylori.

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.

XX 01-APR-1996; 96US-00630405.

XX (ASTR) ASTRA AB.

XX Smith D, Berglindh OT, Mellgaerd BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT68137.

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.

XX Claim 61; Page 1285-87; 1481p; English.

XX This sequence represents a H. pylori cytoplasmic protein. The protein may
CC be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analyzed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analyzed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts

XX Sequence 1183 AA;

Alignment Scores:

Prod. No.: 2 45e-82 Length: 1183
Score: 896.50 Matches: 189
Percent Similarity: 74.8% Conservative: 7
Best Local Similarity: 72.1% Mismatches: 15
Query Match: 76.6% Indels: 51
DB: 2 Gaps: 2

US-09-360-685C-26 (1-685) x AAW20884 (1-1183)

QY 1 AAAAATGGCAAAATAGGATTTTCAGCAGGTAAAGCAAAAGCACTGAAAT 60
DB 770 LysaAnG1LySaAnL1ySaPheSerLyValThrGlnAlaYSerSerPheGluAn 789

QY 61 TCCGTTAAAGTGTGATCATCAATCAAAAGGTAAAGGATTAAGTTGATTAATCTCAATCA 120
DB 790 Ser1eLySaPVal1le1leAnG1nLy1eThSaP1yValaPheAn1eAnG1n 809
QY 121 GCGGTATGAGTGGCTAAAGCAAGCGGTATTTGATGAGTGAAGCGGTAGCCAT 180
DB 810 AlaValSerValAla1ySa1eThrG1ySaPheSerG1yValG1uG1nAla1eAn1aP 829
QY 181 CTCAAAAATTTCTCAAAAGGAGCAATGGCCCAACGCTCAAAATTAAGTCAAT 240
DB 830 Leu1ySaAnPheSerLySeG1uG1nLeuAlaG1nAlaG1nLySaAnG1uAphAn 849
QY 241 GCTAGAAAAATCTGAAATATATCAATCCGTTAAGAAATGGTGAATGGACCCACTC 300
DB 850 ThrG1ySaAnSerAla1eUyG1nSerVal1ySaAnG1yValaAnG1yThrLeuVal 869
QY 301 GGTAAATGGTTATCTCAAGCAAGCAAGCAAGCTTTCTTAAATCTTGGACATCAAG 360
DB 870 G1yAnG1yLeuSerLySa1aG1nAlaThrThrLeuSerLySaAnPheSerP1e1ySa 889
QY 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAACATTAATATGAGCTCAAAAC 420
DB 890 LySaG1uLeuAnAla1ySa1eUyG1nSerPheAnSaAnSaAnSaAnG1yLeuG1uAn 909
QY 421 GAACCCATTTATGCTAAAGTTAATMAAAGAAAGCAAGCAAGCTTGAAGAA 480
DB 910 Ser-----ThrG1u 912
QY 481 CCCATTAAGCTCAAGTGTCTMAAAGGTAAATGCAAAATTAAGCACTCAATCAATA 540
DB 913 Pro1eLyThrG1nG1nValAla1ySa1eVal1ySa1eAla1ySa1eAparG1uAnG1n 932
QY 541 GCAAGTGTGTTGGTGTGTGTAAGGCAAGCAAGCGAGC----- 576
DB 933 AlaSerG1yLeuG1ySaPValG1yG1nAla1aSerPheLeu1eUySaRgh1SaP1ySa 952
QY 576 ----- 576
DB 953 ValaPhePheLeuSerLySa1aG1yLeuSerAlaAnG1nSaG1uPPro1eLyThrAlaThr1le 972
QY 577 -----TTCCCTTGAAGAGGCAATGAATGAATGAATGATCTCAAGTAA 618
DB 973 AaPaPheUyG1yG1yProPhePro1eUySaRgh1SaP1ySa1aPhePheUySa 992
QY 619 GTAGGCTTGAAGAAATGCAATGGCTCAAAATTTGATCAATCAAGCGGTA 678
DB 993 ValG1yLeuSerSaG1uG1nLySa1eUyThrG1nLySa1ePheAn1eAnG1nAlaVal 1012
QY 679 TCAGAA 684
DB 1013 SerGlu 1014

RESULT 12

ABU51383
ID ABU51383 standard; protein, 389 AA.

XX ABU51383;

XX 07-MAY-2003 (first entry)

XX Helicobacter pylori selected interacting domain (SID) protein #727.

XX Protein-protein interaction; ulcer; selected interacting domain; SID.

XX Helicobacter pylori.

XX WO200266501-A2.

XX 29-AUG-2002.

XX 28-DEC-2001; 2001WO-EP015428.

PR 02-JAN-2001; 2001US-0259302P.
 XX (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 XX
 DR WPI; 2002-674910/72.
 DR N-PSDB; ABX66128.
 XX
 PT New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.
 XX
 PS Claim 6; Page 262; 642pp; English.
 XX
 CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (SID), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue
 XX
 SQ Sequence 389 AA;
 Alignment Scores:
 Pred. No.: 3, 01e-67 Length: 389
 Score: 747.50 Matches: 162
 Percent Similarity: 74.2% Conservative: 2
 Best Local Similarity: 73.3% Mismatches: 20
 Query Match: 63.8% Indels: 37
 DB: 5 Gaps: 3
 US-09-360-685C-26 (1-685) x ABUS1383 (1-389)
 QY 130 GTGGCTAAAGCAACGGGATTCAGTAGGAGTAGAGCAGCCGATCCGATCAAAAT 189
 DB 1 ValAlaIyAlaMeRgIyAspPheSerArgValGIuGInValLeuAlaAspLeuIyAsn 20
 QY 190 TTCCTGAAGAGCAATTCGCCCAACAGCTCAAAAATGAAAGTCTCAATCTAGAAAA 249
 DB 21 PheSerIySGluGInLeuAlaGInGInAlaGInIySAsnGIuAspPheAsnThrGIyS 40
 QY 250 AAATCTGAAATATATCAATCCGTTAGAGATGCTGTGAGATGCAACCTTGTGCTAATGGG 309
 DB 41 AsnSerGIuLeuIyGInSerValIySAsnSerValAsnIyVethLeuValGIyAsnIy 60
 QY 310 TTATCTCAAGCAAGCAACACTCTTTCTAAAACTTTTCGAGATCAAGAAAGTTG 369
 DB 61 LeuSerGIyIleGIuAlaThrAlaLeuAlaIySAsnPheSerAspIleIySAsnIy 80
 QY 370 AATGCAAACTTGAAATTCATATCAATCAATTAATGACCTCAAAAC-----GAA 423
 DB 81 AsnGIuIySAsnIySAsnPhe-----AsnAsnAsnAsnAsnGIyLeuIySAsnSerThrGIu 99
 QY 424 CCCATTATGCTTAAAGTTAATTAATAAGAAAGCAAGGCAAGCCTAGCTTGAAGAACCC 483
 DB 100 ProIleIyAlaIyValAsnIySAsnIySAsnIySAsnIySAsnIySAsnIySAsnIy 119
 QY 484 ATTACGCTCAAGTTGCTTAAAGTAAATGCAAAATTTGACCACTCAATCAATATAGCA 543
 DB 120 IleIyThrGInValAlaIySAsnIySAsnIySAsnIySAsnIySAsnIySAsnIy 139
 QY 544 AGTGTGTTGGGTTGTTGAGGAGCAAGCGGCTTCCTTTGAAAGGCAATGATTAAGTT 603
 DB 140 SerGIyLeuGIyGIyValGIyGInAlaIySAsnIySAsnIySAsnIySAsnIySAsnIy 159
 QY 604 GATGATCTCAGTAAGTAGGGCTT----- 627

DB 160 AsnAspLeuSerIyValGIyLeuSerAlaSerProGIuProIleIyAlaThrIleAsp 179
 QY 627 ----- 627
 DB 180 AspLeuGIyGIyProPheProLeuIySAsnIySAsnIySAsnIySAsnIySAsnIy 199
 QY 628 -----TCAGGATTCAGAAATTCGCTCAAGAAATTCGATTCATCAAGCGGTATCA 681
 DB 200 GIyArgSerArgAsnGIuGInLeuAlaGInIySAsnIySAsnIySAsnIySAsnIy 219
 QY 682 GAA 684
 DB 220 GIu 220
 DB 220 GIu 220
 RESULT 13
 ABUS1625
 ID ABUS1625 standard; protein, 356 AA.
 XX
 AC ABUS1625;
 XX
 XX 07-MAY-2003 (first entry)
 XX
 DE Helicobacter pylori selected interacting domain (SID) protein #969.
 XX
 KM Protein-protein interaction; ulcer; selected interacting domain; SID.
 XX
 OS Helicobacter pylori.
 XX
 PN WO20026501-A2.
 XX
 XX 29-AUG-2002.
 PD
 XX 28-DEC-2001; 2001WO-EP015428.
 PF
 XX 02-JAN-2001; 2001US-0259302P.
 PR
 XX (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 XX
 DR WPI; 2002-674910/72.
 DR N-PSDB; ABX66370.
 XX
 PT New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.
 XX
 PS Claim 6; Page 321; 642pp; English.
 XX
 CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (SID), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue
 XX
 SQ Sequence 356 AA;
 Alignment Scores:
 Pred. No.: 7, 22e-63 Length: 356
 Score: 704.50 Matches: 153
 Percent Similarity: 73.5% Conservative: 2
 Best Local Similarity: 72.5% Mismatches: 19
 Query Match: 60.2% Indels: 37
 DB: 5 Gaps: 3

US-09-360-685C-26 (1-685) x ABUS1625 (1-356)

```

QY 160 GTAGAGCAGCGCTTGAAGCAATTTCTCAAGAGCAATGGCCCAACAGCT 219
Db 1 ValGluGlnValLeuAlaAspLeuLysAsnPheserIySgluGlnLeuAlaGlnGlnAla 20
QY 220 CAAAAATGAAAGCTCAATGCTAAGAAAAATTAATATCAATCCGTTAAGAAAT 279
Db 21 GlnLysAsnGluAspPheserIySgluGlnLeuAlaGlnGlnAla 40
QY 280 GGTGTGAATGAAAGCTTGTGCTTAATGGCTTATCTCAAGCAAGCAACCACTTTCT 339
Db 41 SerValAsnLysIleThrLeuValGlnLysGlnLysSerGlyIleGluAlaThrAlaLeuAla 60
QY 340 AAAAAGCTTTCCGACATCAAGAAAGATTGAATGCAAACTTGAATTTCAATTAACAT 399
Db 61 LysAsnPheserIySgluGlnLeuAlaGlnLysGlnLysPheLysAsnPheserIySglu 79
QY 400 AACCAATATGACTCAAAAAC-----GAACCAATTTATGCTTAAGTTAATAAAGAAA 453
Db 80 AsnAsnAsnGlnLysLeuLysAsnSerThrGluProIleTyAlaLysValAsnLysLysLys 99
QY 454 GCAAGGCAAGCAGCTAGCTTGAAGAACCCATTTAGCGTCAAGTTGCTAAAGGTAAT 513
Db 100 ThrGluGlnValAlaSerProGluGluProIleTyThrGlnValAlaLysLysValAsn 119
QY 514 GCAAAAATTTGACGACTCAATCAATATGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 573
Db 120 AlaLysIleAspArgLeuAsnGlnIleLysSerGlyLeuGlnLysGlnLysAla 139
QY 574 GGGTTCCTTTGAAAAGCATGATTAAGTTGATGATCTCAAGTGAAGGAGGCTT----- 627
Db 140 GlyPheProLeuLysArgHisAspLysValAlaAspAspLeuSerLysValGlyLeuSerAla 159
QY 627 ----- 627
Db 160 SerProGluProIleTyAlaThrIleAspAspLeuGlnGlyProPheProLeuLysArg 179
QY 628 -----TCAGGAATCAAGAAATTTGCTCAG 651
Db 180 HisAspLysValAspAspLeuSerLysValGlyArgSerArgAsnGlnLeuAlaGln 199
QY 652 AAAATGCAATCTCAATCAAGCGTATTCGAA 684
Db 200 LysIleAspAsnLeuAsnGlnAlaValSerGln 210

```

RESULT 14

ABUS0882 standard; protein; 379 AA.

ABUS0882;
 AC ABUS0882;
 XX 07-MAY-2003 (first entry)
 XX Helicobacter pylori selected interacting domain (SID) protein #225.
 DE Helicobacter pylori selected interacting domain (SID) protein #225.
 KM Protein-protein interaction; ulcer; selected interacting domain; SID.
 OS Helicobacter pylori.
 XX
 PN WO200266501-A2.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-DEC-2001; 2001WO-EP015428.
 XX
 PR 02-JAN-2001; 2001US-0259302P.
 XX
 PA (HYBR-) HYBRIGENICS.
 XX (INSP) INST PASTEUR.
 XX
 PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

DR MPI: 2002-674910/72.
 DR N-PSDB; ABX65626.

PT New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.

PS Claim 6; Page 140; 642pp; English.

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein.
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (SID), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the index has replaced the residue with an X to
 CC represent an illegible residue

Sequence 379 AA;

Alignment Scores:
 Pred. No.: 7,346-63 Length: 379
 Score: 704.50 Matches: 153
 Percent Similarity: 73.5% Mismatches: 2
 Best Local Similarity: 72.5% Indels: 37
 Query Match: 60.2% Gaps: 3

US-09-360-685C-26 (1-685) x ABUS0882 (1-379)

```

QY 160 GTAGAGCAGCGCTTGAAGCAATTTCTCAAGAGCAATGGCCCAACAGCT 219
Db 1 ValGluGlnValLeuAlaAspLeuLysAsnPheserIySgluGlnLeuAlaGlnGlnAla 20
QY 220 CAAAAATGAAAGCTCAATGCTAAGAAAAATTAATATCAATCCGTTAAGAAAT 279
Db 21 GlnLysAsnGluAspPheserIySgluGlnLeuAlaGlnGlnAla 40
QY 280 GGTGTGAATGAAAGCTTGTGCTTAATGGCTTATCTCAAGCAAGCAACCACTTTCT 339
Db 41 SerValAsnLysIleThrLeuValGlnLysGlnLysSerGlyIleGluAlaThrAlaLeuAla 60
QY 340 AAAAAGCTTTCCGACATCAAGAAAGATTGAATGCAAACTTGAATTTCAATTAACAT 399
Db 61 LysAsnPheserIySgluGlnLeuAlaGlnLysGlnLysPheLysAsnPheserIySglu 79
QY 400 AACCAATATGACTCAAAAAC-----GAACCAATTTATGCTTAAGTTAATAAAGAAA 453
Db 80 AsnAsnAsnGlnLysLeuLysAsnSerThrGluProIleTyAlaLysValAsnLysLysLys 99
QY 454 GCAAGGCAAGCAGCTAGCTTGAAGAACCCATTTAGCGTCAAGTTGCTAAAGGTAAT 513
Db 100 ThrGluGlnValAlaSerProGluGluProIleTyThrGlnValAlaLysLysValAsn 119
QY 514 GCAAAAATTTGACGACTCAATCAATATGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 573
Db 120 AlaLysIleAspArgLeuAsnGlnIleLysSerGlyLeuGlnLysGlnLysAla 139
QY 574 GGGTTCCTTTGAAAAGCATGATTAAGTTGATGATCTCAAGTGAAGGAGGCTT----- 627
Db 140 GlyPheProLeuLysArgHisAspLysValAlaAspAspLeuSerLysValGlyLeuSerAla 159
QY 627 ----- 627
Db 160 SerProGluProIleTyAlaThrIleAspAspLeuGlnGlyProPheProLeuLysArg 179
QY 628 -----TCAGGAATCAAGAAATTTGCTCAG 651
Db 180 HisAspLysValAspAspLeuSerLysValGlyArgSerArgAsnGlnLeuAlaGln 199

```


This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:29:20 ; Search time 3706 Seconds

(without alignments)
8647.909 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685

Sequence: 1 aaaaatgcaacaataagga.....caatcaagcggtatcagaag 685

Scoring table: IDENTITY_NIC

Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.4	8.8	1362	10	AG429915 Mus muscu
2	59.8	8.7	747	10	AL100640 Drosophila
3	59	8.6	1132	9	CC248307 CH261-881
4	58.6	8.6	1007	11	CNS06K9S
5	57.6	8.4	1390	10	AG3447224
6	57.2	8.4	1247	8	DN698953
7	57	8.3	954	5	BX362982
8	56.2	8.2	1239	10	AG305325
9	56.2	8.2	1366	8	DN718180
10	55.6	8.1	1400	10	AG430154
11	54.8	8.0	1364	8	DN693282
12	54.6	8.0	1071	1	AJ926560
13	54.6	8.0	1128	1	CG753708
14	54.4	7.9	1263	10	CL103754
15	54.4	7.9	1266	10	AJ859745
16	54.4	7.9	1475	10	CL495869
17	54.2	7.9	1104	1	AJ925669
18	54	7.9	1435	10	AG3447018
19	53.8	7.9	802	10	AJ885252
20	53.8	7.9	971	6	CD388253
21	53.8	7.9	1101	10	CNS00DG1
22	53.6	7.8	1028	8	DN564006

23	53.6	7.8	1101	10	CNS0182P
24	53.6	7.8	1399	10	AG311276
25	53.4	7.8	907	10	CNS021U4
26	53.4	7.8	976	10	AG136173
27	53	7.7	1137	10	CL080345
28	52.8	7.7	1335	10	CL647352
29	52.8	7.7	1928	8	DR154244
30	52.6	7.7	834	9	B12387
31	52.6	7.7	1101	10	CNS017KX
32	52.6	7.7	1401	10	AG346504
33	52.2	7.6	783	10	CNS011PP
34	52.2	7.6	1173	1	AJ926089
35	52	7.6	1330	8	DN705530
36	52	7.6	1640	10	AG393452
37	51.8	7.6	1195	8	DR146887
38	51.8	7.6	1201	9	BZ570849
39	51.8	7.6	1221	9	CC301561
40	51.8	7.6	1683	10	CL075415
41	51.6	7.5	804	11	CNS03C75
42	51.4	7.5	661	9	AQ378322
43	51.4	7.5	1227	10	CG746198
44	51.4	7.5	1269	10	CG749445
45	51.4	7.5	1291	10	AG387132

ALIGNMENTS

RESULT 1

AG429915/c 1362 bp DNA linear GSS 21-DEC-2004

LOCUS Mus musculus molossinus DNA, Clone:MSMg01-304L12.TU, genomic survey

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57466"
/clone="MSNG01-30412.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSNG01 Mouse Male BAC Library"

ORIGIN

Query Match 8.8%; Score 60.4; DB 10; Length 1362;
Best Local Similarity 36.6%; Pred. No. 0.00052;
Matches 199; Conservative 0; Mismatches 345; Indels 0; Gaps 0;

```

QY 1 AAAAAATGCAAAATTAAGATTTCAGCAAGTAAACCAAGCAAAAGCACTTGAAT 60
DB 1140 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 1081
QY 61 TCCGTTAAAGATGATCATCATCAAAAGTAAAGATTAAGTATATTCATCA 120
DB 1080 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1021
QY 121 GCGGTATCAGTGCCTAAAGCAAGCGGTGATTTCAGTGGGTAGAGCAAGGTTAGCCAT 180
DB 1020 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 961
QY 181 CTCAAAATTTTTCAAAGAGCAATGGCCCAAGCAAGCTCAAAAGTAAAGTCTCAT 240
DB 960 NNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 901
QY 241 GCTGAAAAAATCTGAATATATCATCGTTAAGATGATGATGATGATGATGATG 300
DB 900 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 841
QY 301 GGTATGGGTATCTCAAGAGACCACTCTTTCTTAAAACTTTTGACATCAAG 360
DB 840 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 781
QY 361 AAGAGTTGAATGCAAACTTGAAATTTCAATPACATTAATGATCTCAAAAC 420
DB 780 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 721
QY 421 GAACCATTTTATGATTAATTAATAAAGAAAGCGAGAGCTGCTTGAAGA 480
DB 720 NNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAAN 661
QY 481 CCATTTTACGCTCAAGTTGCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 540
DB 660 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 601
QY 541 GCAA 544
DB 600 AAAA 597

```

RESULT 2
LOCUS CDS011RQ
DEFINITION CDS011RQ 747 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SFE end of BAC
BACN06A03 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL100640
VERSION AL100640.1 GI:5612251
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 747)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

COMMENT

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CSH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
peloBAC11.
Location/Qualifiers
1..747

FEATURES
Source

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN06A03"
/clone_lib="DrosBAC"
/plasmid="peloBAC11"
/note="end : SP6"

ORIGIN

Query Match 8.7%; Score 59.8; DB 10; Length 747;
Best Local Similarity 41.0%; Pred. No. 0.00065;
Matches 128; Conservative 35; Mismatches 149; Indels 0; Gaps 0;

```

QY 156 TAGGGTAAAGCAAGCGTACCGATCTCAAAATTTTTCAGAGCAATGGCCCAAG 215
DB 162 TTTAAATTTTCAAAAGCAAGCGTACCGATCTCAAAATTTTTCAGAGCAATGGCCCAAG 221
QY 216 AACTCAAAATTTTTCAGATCTCAATGCTGAAATTTTTCAGATCTCAATGCTGAA 275
DB 222 AATTTTCAAAATTTTTCAGATCTCAATGCTGAAATTTTTCAGATCTCAATGCTGAA 281
QY 276 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 335
DB 282 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 341
QY 336 TTTTCAAAATTTTTCAGATCTCAAAAGCAAGCGTACCGATCTCAAAATTTTTCAG 395
DB 342 TTTTCAAAATTTTTCAGATCTCAAAAGCAAGCGTACCGATCTCAAAATTTTTCAG 401
QY 396 CATTAATTAATGATCTCAAAAGCAAGCGTACCGATCTCAAAATTTTTCAGATCTCA 455
DB 402 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 461
QY 456 AGGCAAGCAAGC 467
DB 462 AKCAATTCAC 473

```

RESULT 3
LOCUS CC248307/c
DEFINITION CC248307 1132 bp DNA linear GSS 12-MAY-2003
genomic survey sequence.
ACCESSION CC248307
VERSION CC248307.1 GI:30574970
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Krenitzki, C., Higinbotham, T., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 182000 Std Error: 0.00

Seq primer: RM1 TACGACTCACTATAGGAGAGA
 Class: BAC ends
 High quality sequence start: 28
 High quality sequence stop: 165.
 Location/Qualifiers
 1. 1132

/organism="Gallus gallus"
 /mol_type="genomic DNA"
 /strain="Red Jungle Fowl"
 /db_xref="taxon:9031"
 /clone="CH261-8817"
 /sex="female"
 /cell_line="UCD001, inbred 256"
 /clone_11b="CH261"
 /note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
 CH261 Female Chicken library - for library and clone
 ordering information: http://www.choil.org/bacpac"

ORIGIN

Query Match 8.6%; Score 59; DB 9; Length 1132;
 Best Local Similarity 37.3%; Pred. No. 0.0011; Indels 0; Gaps 0;
 Matches 203; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

```

QY 1 AAAAAATGCGCAAAATAGGATTTCAAGCAGTACGCAAGCAAAAGCACTTGAAT 60
DB 1015 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 956
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAGGTAACGATTAAGTTGATCAATCA 120
DB 955 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 896
QY 121 GCGGTATCAGTGCTAAAGCAGCGGTGATTTCAAGGTAGAGCAAGGTTAGCCGAT 180
DB 895 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 836
QY 181 CTCGAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAATATGAACTCAAT 240
DB 835 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 776
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGCTGTAAGTGAACCTAGTC 300
DB 775 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 716
QY 301 GGTATGCGGTATCTCAAGAGCAAGCAACTCTTCTTAAATCTTTCGACATCAAG 360
DB 715 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 656
QY 361 AAGAGTTGATGCAAACTTGGAAATTTCAATTAACAATTAACAATTAACAATTA 420
DB 655 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 596
QY 421 GAACCAATTTTATGCTAAAGTAAAGAAAGAGGCAAGCAGCTGACCTGAAGA 480
DB 595 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 536
QY 481 CCCATTAGCGCTCAAGTGTCTAAAGGTAATGCAAAATTAAGCACTCAATCAATA 540
DB 535 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 476
QY 541 GCNA 544
DB 475 AAAA 472

```

RESULT 4
 CDS06X9S 1007 bp DNA linear GSS 06-UTL-2001
 LOCUS T3 end of clone AX0AA039F08 of library AX0A from strain CBS 7064
 DEFINITION of Pichia farinosa, genomic survey sequence.
 ACCESSION AL419462
 VERSION AL419462.1 GI:12202640
 KEYWORDS GSS.
 SOURCE Pichia farinosa

ORGANISM

Pichia farinosa

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE
 1 (bases 1 to 1007)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaroglia,S.,
 de Montigny,J., Dujon,B., Durrens,P., Legingle,A., Llorente,B.,
 Malpertuy,A., Neveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE

JOURNAL

PUBMED

11152876

2 (bases 1 to 1007)

REFERENCE

de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,
 Wincker,P., Artiguenave,F. and Potier,S.
 Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
 sorbitophila
 FEBS Lett. 487 (1), 87-90 (2000)

JOURNAL

PUBMED

11152890

3 (bases 1 to 1007)

JOURNAL

COMMENT

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxi, Saccharomyces
 kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 fragilis var. fragilis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES

source

1..1007
 /organism="Pichia farinosa"
 /mol_type="genomic DNA"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0AA039F08"
 /clone_11b="AX0AA"
 /note="end : T3"

ORIGIN

Query Match 8.6%; Score 58.6; DB 11; Length 1007;
 Best Local Similarity 37.1%; Pred. No. 0.0014;
 Matches 209; Conservative 62; Mismatches 286; Indels 6; Gaps 1;

```

QY 1 AAAAAATGCGCAAAATAGGATTTCAAGCAGTACGCAAGCAAAAGCACTTGAAT 60
DB 105 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 164
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAGGTAACGATTAAGTTGATCAATCA 120
DB 165 NAAAAAAAAATTAATTTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 224
QY 121 GCGGTATCAGTGCTAAAGCAAGCGGTGATTTCAAGGTAGAGCAAGGTTAGCCGAT 180
DB 225 GTRCATGTGTCWGGTGATGAGWGTATATWAGMAAARARATWAGWTAAGWMAW 284
QY 181 CTCGAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAATATGAAAGTCAAT 240
DB 285 AAAAAAAAAAGAGAGGTTATTAATAAATAAATAAATAAATAAATAAATAAATAA 344
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAA-----GAATGCTGTGAATGAACC 294
DB 345 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404
QY 295 CTAGTCGGTATGCGGTATCTCAAGCAGAGCAACTCTTTCTTAAAACTTTTCGAGC 354

```


/clone.lib="SHGC-CNB"
 /note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand primer: 5'-GACTAGTCTAGATCGGAGCGGCCGCTT25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5' prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxATC' (where x is A/C/G/T). A map of the Express 1 vector is available at:
http://www.openbiosystems.com/cdna_library_construction_fa_q.php#8
 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:
<http://www.openbiosystems.com/stickleback>"

ORIGIN

Query Match 8.2%; Score 56.2; DB 8; Length 1366;
 Best Local Similarity 31.4%; Pred. No. 0.0057;
 Matches 175; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

```

QY 13 AATAGAGATTTCAGCAAGTAACGCAAGAAAGGACCTTGAAATTCGTTAAAGAT 72
DB 1360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1301
QY 73 GGTATCATCAATCAAGTAACGATTAAGTTGATTAATCTCAATACCGGTATCAGTG 132
DB 1300 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1241
QY 133 GCTAAGCAACGCGTGAATTCAGTAGGTAAGCAAGCGTACCGATCTCAAAATTTTC 192
DB 1240 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1181
QY 193 TCAGAGAGCAATTTGGCCCAAGCAAGCTCAAAATGAAAGTCTCAATGCTGAAAAAAA 252
DB 1180 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1121
QY 253 TCTGAATATATCAATCCCTTAAGATGCTGATGGAACCTAGTCGTAATGGGTTA 312
DB 1120 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1061
QY 313 TCTCAGAGCAAGCAAGCACTCTTCTAATAAATCTTTCGACATCAAGAGATTGAAT 372
DB 1060 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1001
QY 373 GCAAACTCTGGAATTTCAATTAACAATTAATGAATCAAAAGCAAGCCATTAT 432
DB 1000 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 941
QY 433 GCTAAGTTAATTAAGAAAGAGAGGCAAGCTGAGCTTGAAGAACCTTTACGCT 492
DB 940 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 881
QY 493 CAAGTTGCTAAAGGTAATGCAAAATTTGACCGATCAATCAATGCAAGTGTGTTG 552
DB 880 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 821
QY 553 GGTGTTGTAGGCAAGCA 570
DB 820 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

```

RESULT 10
 AG430154/c 1400 bp DNA linear GSS 21-DEC-2004
 LOCUS AG430154/c
 DEFINITION Mus musculus molossinus DNA, clone:MSMg01-305A24.TU, genomic survey sequence.

ACCESSION AG430154
 VERSION AG430154.1 GI:48073217
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)
 ORGANISM Mus musculus molossinus
 Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 COMMENT
 1 Abe, K., Noguichi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K. and Shiroishi, T.
 Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
 Genome Res. 14 (12), 2439-2447 (2004)
 2 (bases 1 to 1400)
 Hattori, M., Toyoda, A., Noguichi, H., Kojima, T. and Sakaki, Y.
 Direct Submission
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: hattori@gsc.riken.jp URL: <http://ngp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS
 Sequencing : TJ
 LIBRARY
 Vector : pBAC3.6
 R.site 1 : EcoRI
 R.site 2 : EcoRI

FEATURES

source
 1..1400
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-305A24.TU"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 8.1%; Score 55.6; DB 10; Length 1400;
 Best Local Similarity 29.7%; Pred. No. 0.0081;
 Matches 160; Conservative 0; Mismatches 378; Indels 0; Gaps 0;

```

QY 1 AAAAAAGCAAAATTAAGATTTCAGCAAGTAACGCAAGAAAGGACCTTGAAAT 60
DB 1175 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1116
QY 61 TCCGTTAAGTGTGATCAATCAATCAAAAGTAAGATTAAGTTAATCTCAATCAA 120
DB 1115 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1056
QY 121 GCGGTATCAGTGTGCTAAGCAACGCGTGAATTCAGTAGGTAAGAGCGTTAGCGAT 180
DB 1055 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 996
QY 181 CTCAAAATTTCTCAAGAGCAATTTGCCCAAGCAAGCTCAAAATTAAGATCTCAAT 240
DB 995 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 936
QY 241 GCTGAAGAAATATCTGAATATCAATCCGTTAAGATGTGTGAATGGAACCTTAGTC 300
DB 935 NNNNAANNNNNNNNNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNN 876

```


QY 301 GGTAAATGGCTTATCTCAAGCAAGCCCACTCTTCTAAACCTTTGCGACATCAAG 360
 Db 875 NNN 816
 QY 361 AAGAGTTGATGCAAACTGGAAATTTCAATTAACAATTAACAATTAACAATTAACA 420
 Db 815 NNN 756
 QY 421 GAACCCATTTTGTCTAAAGTTAATTAAGAAAGAGGCGACGACGCTGCTGAAGA 480
 Db 755 NNN 696
 QY 481 CCCATTTCGCTCAAGTTGCTAAAGTTAATGCAAAATTTGACCGCATCATCAAA 538
 Db 695 NNN 638

RESULT 11
 DNE93282 1364 bp mRNA linear EST 30-MAR-2005
 LOCUS DNE93282
 DEFINITION CC887-B03.3'-c SHGC-CGX Gaetereosteus aculeatus cDNA clone
 ACCESSION DNE93282
 VERSION DNE93282.1 GI:62049511
 KEYWORDS EST.
 SOURCE Gaetereosteus aculeatus (three spined stickleback)
 ORGANISM Gaetereosteus aculeatus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Gaetereostei;
 Gaetereosteiidae; Gaetereosteus.
 1 (bases 1 to 1364)
 Kinsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
 Schmutz,J. and Myers,R.M.
 Unpublished sequence tags from Gaetereosteus aculeatus
 (2003)
 CONTACT: Grimwood, Jane
 Stanford Human Genome Center
 Stanford University School of Medicine
 975 S California Ave, Palo Alto, CA 94304, USA
 Tel: 650 320 5917
 Fax: 650 320 5801
 Email: jane@shgc.stanford.edu
 Plate: 87

REFERENCE
 JOURNAL
 COMMENT
 FEATURES
 source
 High quality sequence start: 112
 High quality sequence stop: 368.
 Location/Qualifiers
 1..1364
 /organism="Gaetereosteus aculeatus"
 /mol_type="mRNA"
 /strain="Conner Creek sticklebacks, WA USA"
 /db_xref="taxon:69293"
 /clone="CGX87-B03"
 /sex="mixed male and female"
 /tissue_type="eyes"
 /dev_stage="adult"
 /lab_host="DH10B (T1 phage resistant)"
 /note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by with an 54 bp linker primer containing an oligodT sequence preceded by a synthetic NotI site (first strand synthesis 5'-GACAGCTTAACTGCGACGCGCCCTT) (25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of the Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of

the Express 1 vector is available at:
http://www.openbioystems.com/cdna_library_construction_fa
 g.php#6 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Bioystems:
<http://www.openbioystems.com/stickleback>

Query Match 8.0%; Score 54.8; DB 8; Length 1364;
 Best Local Similarity 35.8%; Pred. No. 0.013;
 Matches 195; Conservative 0; Mismatches 346; Indels 3; Gaps 1;
 QY 1 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 60
 Db 592 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 60
 QY 61 TCCGTTAAAGTGTGATCATCAATCAAAAGTAAAGTAAAGTAAATCTGATCA 120
 Db 652 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 120
 QY 121 GCGGTATGATGCTTAAGCAACGGGTGATTTGATGAGGTGAGCAAGCGCTTACCGCAT 180
 Db 712 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 180
 QY 181 CTCGAAATTTCTCAAGAGCAATTGGCCCAAGCTCAAAATTAAGTCAAT 240
 Db 772 NNN 240
 QY 241 GCTGCAAAATTAAGATTTGATCAATCGTTAAGATGATGATGATGATGATGATGATG 300
 Db 832 NNN 300
 QY 301 GGTAAATGGCTTATCTCAAGCAAGCCCACTCTTCTAAACCTTTGCGACATCAAG 360
 Db 889 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 360
 QY 361 AAGAGTTGATGCAAACTGGAAATTTCAATTAACAATTAACAATTAACAATTAACA 420
 Db 949 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 420
 QY 421 GAACCCATTTTGTCTAAAGTTAATTAAGAAAGAGGCGACGACGCTGCTGAAGA 480
 Db 1009 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 480
 QY 481 CCCATTTCGCTCAAGTTGCTAAAGTTAATGCAAAATTTGACCGCATCATCAAA 538
 Db 1069 AAAAAATGCAAAATTAAGATTTGACGAGGTAAGCGCAAGCAAAAGCGACTGAAAT 538
 QY 541 GCNA 544
 Db 1129 NAAA 1132

RESULT 12
 DNE926560 1071 bp mRNA linear EST 10-JUN-2005
 LOCUS DNE926560
 DEFINITION DNE926560 Theileria annulata metazoite Theileria annulata cDNA clone
 ACCESSION DNE926560
 VERSION DNE926560.1 GI:67496941
 KEYWORDS EST.
 SOURCE Theileria annulata
 ORGANISM Theileria annulata
 Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileridae;
 Theileriinae;
 1 (bases 1 to 1071)
 Pain,A., Renaud,H., Berriman,M., Murphy,L., Yeates,C.A., Weir,W.,
 Kesteven,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
 Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Foster,N.,
 Lakke,N., Lord,A., Griffiths-Jones,S., Harris,D.E., Kitzer,F.,
 Nene,V., O'Neill,S., Price,C., Quail,M.A., Mooney,P., Morton,F.,
 Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,

TITLE
Squares, R., Squares, S., Tivey, A., Walker, A.R., Woodward, J.,
Dobbelare, D.A.E., Langsley, G., Rajandream, M.-A., McKeever, D.,
Shiels, B., Tait, A., Bartlett, B. and Hall, N.
The genome of the host-cell transforming parasite *Theileria*
annulata and a comparison with *T. parva*
Unpublished (2005)

JOURNAL
Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
Location/Qualifiers
1. 1071
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="Cam015f11.g1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey; Ankara"

FEATURES

source

ORIGIN

Query Match 8.0%; Score 54.6; DB 1; Length 1071;
Best Local Similarity 35.1%; Pred. No. 0.014;
Matches 159; Conservative 0; Mismatches 294; Indels 0; Gaps 0;

QY 1 AAAAATGCGAATAATAGATTTCAGCAAGTAAACCAAGCAAAAGCGACTGTAAT 60
DB 586 AA 645
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGTAAATCTCAATCAA 120
DB 646 AA 705
QY 121 GCGGTATCAGTGGCTAAAGCAAGCGGTGATTGAGTGAGTAAAGCAAGCTTACCGCAT 180
DB 706 NAANNNAANNNAANNNAANNNAANNNAANNNAANNNAANNNAANNNAANNNA 765
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCTCAAT 240
DB 766 NNAA 825
QY 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGTGTGAATGAAACCTTAGTC 300
DB 826 AAAAAAAAAAAAAAAAAAAAAAAAAATNANNNAANNNAANNNAANNNAANNNA 885
QY 301 GGTATGGTATCTCAAGCAAGCAAGCACTCTTTCTAATAACTTTTCGACATCAAG 360
DB 886 NNAAAAAAAAAAAAAAAAAAAAAAAAAGAAATATTAATAAAAAAAAAAANGAAGAAAAA 945
QY 361 AAAGAGTTAATGCAAACTTGAATTTCAATTAACATTAACATTAATGAGCTCAAAAC 420
DB 946 AAAAAAAAAAAAAAAAAAATGATTAATGAAATTAATAATNNAANNNAATTAATTA 1005
QY 421 GAACCAATTTATGCTAAAGTAAATAAAGAAA 453
DB 1006 AAAAAAAAAATTATATAAAGAGAAAAAGATATA 1038

RESULT 13
CG753708/c 1128 bp DNA linear GSS 24-OCT-2003
LOCUS P048-4-F03_Ya Ppa EcORI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG753708
VERSION CG753708.1 GI:37978462
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

REFERENCE
1 (bases 1 to 1128)
Srinivasan, V., Simz, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K.,
Buntjer, J., van der Meulen, M. and Sommer, R.J.
An integrated physical and genetic map of the nematode *Pristionchus*
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)

JOURNAL
Contact: Sommer R
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1. 1128
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcORI BAC library"
/note="The library was generated by a partial digest of
the genomic DNA with EcORI and cloning into the BAC
vector."

FEATURES

source

ORIGIN

Query Match 8.0%; Score 54.6; DB 10; Length 1128;
Best Local Similarity 40.8%; Pred. No. 0.014;
Matches 222; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 1 AAAAATGCGAATAATAGATTTCAGCAAGTAAAGCAAAAGCGACTGTAAT 60
DB 996 AA 937
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGTAAATCTCAATCAA 120
DB 936 AA 877
QY 121 GCGGTATCAGTGGCTAAAGCAAGCGGTGATTGAGTGAGTAAAGCAAGCTTACCGCAT 180
DB 876 AA 817
QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCTCAAT 240
DB 816 NNAA 757
QY 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAGATGTGTGAATGAAACCTTAGTC 300
DB 756 AAAAAAAAAAAAAAAAAAAAAAAAAATNANNNAANNNAANNNAANNNAANNNA 697
QY 301 GGTATGGTATCTCAAGCAAGCAAGCACTCTTTCTAATAACTTTTCGACATCAAG 360
DB 696 NNAAAAAAAAAAAAAAAAAAAAAAAAAGAAATATTAATAAAAAAAAAAANGAAGAAAAA 637
QY 361 AAAGAGTTAATGCAAACTTGAATTTCAATTAACATTAACATTAATGAGCTCAAAAC 420
DB 636 AAAAAAAAAAAAAAAAAAATGATTAATGAAATTAATAATNNAANNNAATTAATTA 577
QY 421 GAACCAATTTATGCTAAAGTAAATAAAGAGCAAGCAAGCAAGTAAAGCA 480
DB 576 AAAAAAAAAAAAAAAAAAANGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 517
QY 481 CCCATTACGCTCAAGTGTCTAATAAAGTAAATGCAAAATTTGACCGACTCAATTA 540
DB 516 AAAAAAAAAAAAAAAAAAANGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 457
QY 541 GCAA 544
DB 456 AAAA 453

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 6, 2006, 22:11:09 ; Search time 351 Seconds
(without alignments)
4278.782 Million cell updates/sec

Title: US-09-360-685C-26

Sequence: 1 aaaaatgcaaaaataagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC
Gapop 10.0, Gapext 1.0

Searched: 7218535 seqs, 1096242582 residues

Total number of hits satisfying chosen parameters: 14437070

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database: Published Applications NA New:*

1: /cgn2_6/ptodata/2/pubphn/US08_NEW_PUB.seq:*
2: /cgn2_6/ptodata/2/pubphn/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubphn/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubphn/PCT_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubphn/US09_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubphn/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubphn/US10_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubphn/US10_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubphn/US11_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubphn/US11_NEW_PUB.seq:*
11: /cgn2_6/ptodata/2/pubphn/US11_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubphn/US11_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubphn/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the total score distribution.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	685	100.0	5925 8 US-10-615-668-4	Sequence 4, Appl1
2	477.2	69.7	3504 12 US-11-052-554A-505	Sequence 505, App
3	48.2	7.0	571 6 US-09-925-065A-20908	Sequence 20908, A
4	46.2	6.7	3143 12 US-11-165-819-9	Sequence 9, Appl1
5	46.2	6.7	3143 12 US-11-165-819-29	Sequence 29, Appl1
6	45	6.6	45 8 US-10-615-668-22	Sequence 22, Appl1
7	43.4	6.3	19233 8 US-10-240-708-45	Sequence 45, Appl1
8	43	6.3	648 6 US-09-925-065A-684219	Sequence 684219, A
9	42.2	6.2	533 6 US-09-925-065A-127945	Sequence 127945, A
10	42.2	6.2	633 6 US-09-925-065A-140562	Sequence 140562, A
11	41.2	6.0	524 6 US-09-925-065A-575490	Sequence 575490, A
12	40.6	5.9	10467 8 US-10-240-708-1	Sequence 1, Appl1
13	39.8	5.8	598 6 US-09-925-065A-410917	Sequence 410917, A
14	39.8	5.8	670 6 US-09-925-065A-474252	Sequence 474252, A
15	39.8	5.8	670 6 US-09-925-065A-474253	Sequence 474253, A
16	39.6	5.8	102 8 US-10-615-668-11	Sequence 11, Appl1
17	39.6	5.8	3221 8 US-10-793-626-4163	Sequence 4163, Ap
18	39.6	5.8	3221 8 US-10-793-626-3355	Sequence 3355, Ap
19	39.2	5.7	539 6 US-09-925-065A-237166	Sequence 237166, A
20	39.2	5.7	605 6 US-09-925-065A-524453	Sequence 524453, A

21	39.2	5.7	1488 6 US-09-925-065A-51174	Sequence 51174, A
22	39.2	5.7	1488 6 US-09-925-065A-51175	Sequence 51175, A
23	39.2	5.7	1488 6 US-09-925-065A-51176	Sequence 51176, A
24	39	5.7	2407 8 US-10-485-517-42	Sequence 42, Appl1
25	39	5.7	9347 8 US-10-240-708-35	Sequence 35, Appl1
26	38.8	5.7	605 6 US-09-925-065A-524454	Sequence 524454, A
27	38.6	5.6	45 8 US-10-615-668-18	Sequence 18, Appl1
28	38.6	5.6	487 6 US-09-925-065A-104585	Sequence 304585, A
29	38.6	5.6	539 6 US-09-925-065A-328934	Sequence 328934, A
30	38.6	5.6	539 6 US-09-925-065A-328935	Sequence 328935, A
31	38.6	5.6	568 6 US-09-925-065A-107617	Sequence 107617, A
32	38.6	5.6	598 6 US-09-925-065A-410918	Sequence 410918, A
33	38.6	5.6	179597 12 US-11-121-086-91	Sequence 91, Appl1
34	38.4	5.6	622 6 US-09-925-065A-873655	Sequence 873655, A
35	38.4	5.6	644 6 US-09-925-065A-418161	Sequence 418161, A
36	38.4	5.6	1290 6 US-09-925-065A-705891	Sequence 705891, A
37	38.4	5.6	10144 8 US-10-240-708-93	Sequence 93, Appl1
38	38.4	5.6	1131 8 US-10-240-708-27	Sequence 27, Appl1
39	38.4	5.5	842 8 US-10-750-185-42851	Sequence 42851, A
40	38	5.5	842 8 US-10-750-623-42851	Sequence 42851, A
41	38	5.5	842 8 US-10-750-623-42851	Sequence 42851, A
42	38	5.5	954 8 US-10-750-185-60688	Sequence 60688, A
43	38	5.5	954 8 US-10-750-623-60688	Sequence 60688, A
44	38	5.5	139054 12 US-11-121-086-96	Sequence 96, Appl1
45	37.8	5.5	631 6 US-09-925-065A-215960	Sequence 215960, A

ALIGNMENTS

RESULT 1	US-10-615-668-4	Sequence 4, Application US/10615668
1	US-10-615-668-4	Sequence 4, Application US/10615668
2	US-10-615-668-4	Sequence 4, Application US/10615668
3	US-10-615-668-4	Sequence 4, Application US/10615668
4	US-10-615-668-4	Sequence 4, Application US/10615668
5	US-10-615-668-4	Sequence 4, Application US/10615668
6	US-10-615-668-4	Sequence 4, Application US/10615668
7	US-10-615-668-4	Sequence 4, Application US/10615668
8	US-10-615-668-4	Sequence 4, Application US/10615668
9	US-10-615-668-4	Sequence 4, Application US/10615668
10	US-10-615-668-4	Sequence 4, Application US/10615668
11	US-10-615-668-4	Sequence 4, Application US/10615668
12	US-10-615-668-4	Sequence 4, Application US/10615668
13	US-10-615-668-4	Sequence 4, Application US/10615668
14	US-10-615-668-4	Sequence 4, Application US/10615668
15	US-10-615-668-4	Sequence 4, Application US/10615668
16	US-10-615-668-4	Sequence 4, Application US/10615668
17	US-10-615-668-4	Sequence 4, Application US/10615668
18	US-10-615-668-4	Sequence 4, Application US/10615668
19	US-10-615-668-4	Sequence 4, Application US/10615668
20	US-10-615-668-4	Sequence 4, Application US/10615668
21	US-10-615-668-4	Sequence 4, Application US/10615668
22	US-10-615-668-4	Sequence 4, Application US/10615668
23	US-10-615-668-4	Sequence 4, Application US/10615668
24	US-10-615-668-4	Sequence 4, Application US/10615668
25	US-10-615-668-4	Sequence 4, Application US/10615668
26	US-10-615-668-4	Sequence 4, Application US/10615668
27	US-10-615-668-4	Sequence 4, Application US/10615668
28	US-10-615-668-4	Sequence 4, Application US/10615668
29	US-10-615-668-4	Sequence 4, Application US/10615668
30	US-10-615-668-4	Sequence 4, Application US/10615668
31	US-10-615-668-4	Sequence 4, Application US/10615668
32	US-10-615-668-4	Sequence 4, Application US/10615668
33	US-10-615-668-4	Sequence 4, Application US/10615668
34	US-10-615-668-4	Sequence 4, Application US/10615668
35	US-10-615-668-4	Sequence 4, Application US/10615668
36	US-10-615-668-4	Sequence 4, Application US/10615668
37	US-10-615-668-4	Sequence 4, Application US/10615668
38	US-10-615-668-4	Sequence 4, Application US/10615668
39	US-10-615-668-4	Sequence 4, Application US/10615668
40	US-10-615-668-4	Sequence 4, Application US/10615668
41	US-10-615-668-4	Sequence 4, Application US/10615668
42	US-10-615-668-4	Sequence 4, Application US/10615668
43	US-10-615-668-4	Sequence 4, Application US/10615668
44	US-10-615-668-4	Sequence 4, Application US/10615668
45	US-10-615-668-4	Sequence 4, Application US/10615668
46	US-10-615-668-4	Sequence 4, Application US/10615668
47	US-10-615-668-4	Sequence 4, Application US/10615668
48	US-10-615-668-4	Sequence 4, Application US/10615668
49	US-10-615-668-4	Sequence 4, Application US/10615668
50	US-10-615-668-4	Sequence 4, Application US/10615668
51	US-10-615-668-4	Sequence 4, Application US/10615668
52	US-10-615-668-4	Sequence 4, Application US/10615668
53	US-10-615-668-4	Sequence 4, Application US/10615668
54	US-10-615-668-4	Sequence 4, Application US/10615668
55	US-10-615-668-4	Sequence 4, Application US/10615668
56	US-10-615-668-4	Sequence 4, Application US/10615668
57	US-10-615-668-4	Sequence 4, Application US/10615668
58	US-10-615-668-4	Sequence 4, Application US/10615668
59	US-10-615-668-4	Sequence 4, Application US/10615668
60	US-10-615-668-4	Sequence 4, Application US/10615668
61	US-10-615-668-4	Sequence 4, Application US/10615668
62	US-10-615-668-4	Sequence 4, Application US/10615668
63	US-10-615-668-4	Sequence 4, Application US/10615668
64	US-10-615-668-4	Sequence 4, Application US/10615668
65	US-10-615-668-4	Sequence 4, Application US/10615668
66	US-10-615-668-4	Sequence 4, Application US/10615668
67	US-10-615-668-4	Sequence 4, Application US/10615668
68	US-10-615-668-4	Sequence 4, Application US/10615668
69	US-10-615-668-4	Sequence 4, Application US/10615668
70	US-10-615-668-4	Sequence 4, Application US/10615668
71	US-10-615-668-4	Sequence 4, Application US/10615668
72	US-10-615-668-4	Sequence 4, Application US/10615668
73	US-10-615-668-4	Sequence 4, Application US/10615668
74	US-10-615-668-4	Sequence 4, Application US/10615668
75	US-10-615-668-4	Sequence 4, Application US/10615668
76	US-10-615-668-4	Sequence 4, Application US/10615668
77	US-10-615-668-4	Sequence 4, Application US/10615668
78	US-10-615-668-4	Sequence 4, Application US/10615668
79	US-10-615-668-4	Sequence 4, Application US/10615668
80	US-10-615-668-4	Sequence 4, Application US/10615668
81	US-10-615-668-4	Sequence 4, Application US/10615668
82	US-10-615-668-4	Sequence 4, Application US/10615668
83	US-10-615-668-4	Sequence 4, Application US/10615668
84	US-10-615-668-4	Sequence 4, Application US/10615668
85	US-10-615-668-4	Sequence 4, Application US/10615668
86	US-10-615-668-4	Sequence 4, Application US/10615668
87	US-10-615-668-4	Sequence 4, Application US/10615668
88	US-10-615-668-4	Sequence 4, Application US/10615668
89	US-10-615-668-4	Sequence 4, Application US/10615668
90	US-10-615-668-4	Sequence 4, Application US/10615668
91	US-10-615-668-4	Sequence 4, Application US/10615668
92	US-10-615-668-4	Sequence 4, Application US/10615668
93	US-10-615-668-4	Sequence 4, Application US/10615668
94	US-10-615-668-4	Sequence 4, Application US/10615668
95	US-10-615-668-4	Sequence 4, Application US/10615668
96	US-10-615-668-4	Sequence 4, Application US/10615668
97	US-10-615-668-4	Sequence 4, Application US/10615668
98	US-10-615-668-4	Sequence 4, Application US/10615668
99	US-10-615-668-4	Sequence 4, Application US/10615668
100	US-10-615-668-4	Sequence 4, Application US/10615668

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rippn

Page 2

OY	181	CTCAAAATTTCTCAAAAGACAAATTGGCCCAACAGCTCAAAAAAATGAAGTCAAT	240
Db	2962	CTCAAAATTTCTCAAAAGACAAATTGGCCCAACAGCTCAAAAAAATGAAGTCAAT	3022
OY	241	GCTGAAAAAATCTGAAATTTTCAATCCGTTAAGATGCTGTAATGAAACCTGATC	300
Db	3022	GCTGAAAAAATCTGAAATTTTCAATCCGTTAAGATGCTGTAATGAAACCTGATC	3081
OY	301	GCTAAGGCTATCTCAAGCAAGACCACTCTTTCTAAAACTTTTCGACATCAAG	360
Db	3082	GCTAATGGTATCTCAAGCAAGACCACTCTTTCTAAAACTTTTCGACATCAAG	3141
OY	361	AAAGATGATGCAAAAATTGGAAATTTCAATACATATACATATATGATCACTCAAAAC	420
Db	3142	AAAGATGATGCAAAAATTGGAAATTTCAATACATATATGATCACTCAAAAC	3201
OY	421	GAACCCATTTATCTAAATTAATAAAGAAAGCAGCAGCAAGCTGCTGAAGAA	480
Db	3202	GAACCCATTTATCTAAATTAATAAAGAAAGCAGCAGCAAGCTGCTGAAGAA	3261
OY	481	CCCATTTACGCTCAAGTGGCTAAAAAGATAATGCAAAATTGACAGCTCAATCAAT	540
Db	3262	CCCATTTACGCTCAAGTGGCTAAAAAGATAATGCAAAATTGACAGCTCAATCAAT	3321
OY	541	GCAAGTGTTTGGGTGTGTGAGGCAACAGCCGGCTTCCTTTGAAAAAGCATATATA	600
Db	3322	GCAAGTGTTTGGGTGTGTGAGGCAACAGCCGGCTTCCTTTGAAAAAGCATATATA	3381
OY	601	GTTGATGATCTCAAGTAAGTAAGGAGCTTTCAGAGATATCAAGAAATTGGCTAGAAAAATTGAC	660
Db	3382	GTTGATGATCTCAAGTAAGTAAGGAGAGCTTTCAGAGATATCAAGAAATTGGCTAGAAAAATTGAC	3441
OY	661	AATTCATCAACAGCGATCAAGAG	685
Db	3442	AATTCATCAACAGCGATCAAGAG	3466

```

RESULT 2
US-11-052-554A-505
; Sequence 505, Application US/11052554A
; Publication No., US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKEN
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 505
; LENGTH: 3504
; TYPE: DNA
; ORGANISM: Helicobacter pylori J99
; US-11-052-554A-505

```

Query Match	69.7%	Score 477.2	DB 12	Length 3504
Best Local Similarity	86.9%	Fred. No. 5.3e-100		
Matches	558	Conservative	0	Mismatches 31; Indels 51; Gaps 1;
QY	1	AAAAATGGCAAAATATAGATTTCAGCGAAGTAAAGCGAAGCAAAAGCGACCTTGAAGAT	60	
Db	2260	AAAAATGGCAAAATATAGATTTCAGCGAAGTAAAGCGAAGCAAAAGCGACCTTGAAGAT	2319	
QY	61	TCCGTTAAAGTGTGATCATCATCAAAAGTAAAGCGATTAAGTTATATCTCAATCA	120	
Db	2320	TCCATTTAAAGTGTGATCATCATCAAAAGTAAAGCGATTAAGTTATATCTCAATCA	2379	

Qy	121	GGGATATCATGGCTAAAGCGAAGGGTATTTCTGATGGGTGAGCAAGCTTATAGCCGAT	180
Db	2380	GGGGATCATGGCTTAAGGCAAGGGTATTTCTGAGGGTGAAGCAAGGTTAGCCGAT	2433
Qy	181	CTCAAAAATTTCTCAAAAGGAGCAATTGGCCCAACAAGCTCAAAAAAATGAAGTCTAAT	240
Db	2440	CTCAAAAATTTCTCAAAAGGAGCAATTGGCTCAACAAGCTCAAAAAAATGAAGTTCAT	2493
Qy	241	GCTGAAAAAAAATCTGAATATATCAATCCGTTAAATAGTGTGAATGAAACCTTATGTC	300
Db	2500	ACTGGAAAAAATTTGGCATATACCAATCCGTTAAATAGTGTGAATGAAACCTTATGTC	2555
Qy	301	GGTATAGGGTATCTCAAGAGAGAGACGACATCTTTCTAAAACTTTTGGACATCAAG	360
Db	2560	GGTATAGGGTATCTAAGAGAGACGACATCTTTCTAAAACTTTTGGACATCAAG	2619
Qy	361	AAAAGTTTGAAGCAAACTTGGAATTTTCATTAACATTAACATTAATGATGATCAAAAC	420
Db	2620	AAAAGTTTGAAGCAAACTTGGAATTTTCATTAACATTAACATTAATGATGATGAAAC	2679
Qy	421	GAACCATTTATGCTAAAGTTAATAAAAAAGAAAGCGACGACGTAAGCTTGAAGAA	480
Db	2680	AGCAC-----AGAA	2688
Qy	481	CCCATTTTACGCTCAAGTTGCTAAAAAAGTTAAATGCAAAATTAAGCACTCAATCAATA	540
Db	2688	CCCATTTTAACTCAAGTTGCTAAAAAAGTTAAATGCAAAATTAAGCACTCAATCAATA	2748
Qy	541	GCAAGTGTGTTGGGTGTGTAGAGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAA	600
Db	2749	GCAAGTGTGTTGGGTGTGTAGAGCAAGCAGCGACCTCTTTTGAAGAAGGCATGATAA	2808
Qy	601	GTTGATGATCTCGAAGTAGAGTAGAGCTTTCACAGAAATCAAGAA	642
Db	2809	GTTGATGATCTCGAAGTAGAGTAGAGCTTTCACAGTAACATGAA	2850

```

RESULT 3
US-09-925-065A-20908
? Sequence 20908, Application US/09925065A
? Publication No. US20060181048A1
GENERAL INFORMATION:
? APPLICANT: Wang, David G.
? TITLE OF INVENTION: Identification and Mapping of Single
? TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
? FILE REFERENCE: 10827.135
CURRENT APPLICATION NUMBER: US/09/925, 065A
CURRENT FILING DATE: 2001-08-08
PRIORITY APPLICATION NUMBER: US 60/243, 096
PRIORITY FILING DATE: 2000-10-24
PRIORITY APPLICATION NUMBER: US 60/252, 147
PRIORITY FILING DATE: 2000-11-20
PRIORITY APPLICATION NUMBER: US 60/250, 092
PRIORITY FILING DATE: 2000-11-30
PRIORITY APPLICATION NUMBER: US 60/261, 766
PRIORITY FILING DATE: 2001-01-16
PRIORITY APPLICATION NUMBER: US 60/289, 846
PRIORITY FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 557086
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 20908
LENGTH: 571
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-925-065A-20908

```

Query Match	7.0%	Score 48.2	DB 6	Length 571
Best Local Similarity	50.7%	Pred. No. 0.11		
Matches 116	Conservative 0	Mismatches 113	Indels 0	Gaps 0
Qy	221	AAAAAAAAAGAGTTCATGCTAGAAAAAAATCTGAATATATCATCGTTAGAAAG	280	
Db	24	MAAAAACTCAACGTAAACCCCAATTGAAAAAGGGGCAAAAGCTTCGACGCAATTCAGAAAT	A3	

QY 281 GTGTGATGGAACCTTAGTGGTAATGGTTATCTCAAGAGGACCAACTCTTCTA 340
DB 84 GTATGTTGTATATGATGATTTGTATTCATTAAITGGATCCATGGCCATTATATATG 143
QY 341 AAAATCTTTGCAATCAAGAAAAGTTGAATGCAAACTTGAAATTTCAATACATTA 400
DB 144 AAAAGATATGTGACTTCATCAGCCATTAGGAAAATTAAGATTTAAACTGCAATGAGCTA 203
QY 401 ACAATATGACTCCAAAAGAAACCCATTATGCTAAAGTTAATAAAA 449
DB 204 AACGATATGAAAAAAGAAATATTCCAATGTTAGAAAAGATGTGAAAA 252

RESULT 4
US-11-165-819-9

Sequence 9, Application US/11165819
Publication No. US20060019314A1
GENERAL INFORMATION:
APPLICANT: Ma, Hui
APPLICANT: Dillin, Andrew
APPLICANT: Hunter, Tony
TITLE OF INVENTION: Increasing Life Span by Modulation of
FILE REFERENCE: the Smek Gene
CURRENT APPLICATION NUMBER: US/11/165,819
PRIOR FILING DATE: 2005-06-24
PRIOR APPLICATION NUMBER: 60/583,284
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 3143
TYPE: DNA
ORGANISM: Dictyostelium discoideum
US-11-165-819-9

Query Match 6.7%; Score 46.2; DB 12; Length 3143;
Best Local Similarity 46.7%; Pred. No. 0.52;
Matches 147; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 357 CAGAAAGAGTTGAATGCAAACTTGAAATTTCAATACATTAATGACTCA 416
DB 2643 CGAACTGTTTAAATGTAAGTAATTAATTAATTAATTAATTAATTA 2702
QY 417 AAACGAACCATTTATGCTAAAGTAAATAAAGAAAGCAAGCAAGCTTGA 476
DB 2703 TAATGCAATTAATCAAGATTTGAGATTAATAGAAAACCAACCAAAAGAAATTGA 2762
QY 477 AGAACCCATTTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACCTCATCA 536
DB 2763 TTAATGAAAAAATGAATCTGTTGTTTCAAGAAATTTGAAAGTAAGCAACTTC 2822
QY 537 AATAGCAAGTGTGGTGTGTGTGAGGCAAGCGGCGCTTCCCTTGAAAAAGCATA 596
DB 2823 AATAGCAAAAGATTAATGTTGCGATGATCAACCAATTAATAATTAATTAATTA 2882
QY 597 TAAAGTTGATCTTCAGTAAGTGGGCTTTCAAGAAATCAAGATTTGGCTCAGAAAT 656
DB 2883 TAGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2942
QY 657 TGACAATCTCATCA 671
DB 2943 TAAATTAATTAATTA 2957

RESULT 5
US-11-165-819-29

Sequence 29, Application US/11165819
Publication No. US20060019314A1
GENERAL INFORMATION:
APPLICANT: Ma, Hui
APPLICANT: Dillin, Andrew

APPLICANT: Hunter, Tony
TITLE OF INVENTION: Increasing Life Span by Modulation of
FILE REFERENCE: the Smek Gene
CURRENT APPLICATION NUMBER: US/11/165,819
PRIOR FILING DATE: 2005-06-24
PRIOR APPLICATION NUMBER: 60/583,284
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 3143
TYPE: DNA
ORGANISM: Dictyostelium discoideum
US-11-165-819-29

Query Match 6.7%; Score 46.2; DB 12; Length 3143;
Best Local Similarity 46.7%; Pred. No. 0.52;
Matches 147; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 357 CAGAAAGAGTTGAATGCAAACTTGAAATTTCAATACATTAATGACTCA 416
DB 2643 CGAACTGTTTAAATGTAAGTAATTAATTAATTAATTAATTAATTA 2702
QY 417 AAACGAACCATTTATGCTAAAGTAAATAAAGAAAGCAAGCAAGCTTGA 476
DB 2703 TAATGCAATTAATCAAGATTTGAGATTAATAGAAAACCAACCAAAAGAAATTGA 2762
QY 477 AGAACCCATTTACGCTCAAGTTGCTAAAGTAATGCAAAATTTGACCTCATCA 536
DB 2763 TTAATGAAAAAATGAATCTGTTGTTTCAAGAAATTTGAAAGTAAGCAACTTC 2822
QY 537 AATAGCAAGTGTGGTGTGTGTGAGGCAAGCGGCGCTTCCCTTGAAAAAGCATA 596
DB 2823 AATAGCAAAAGATTAATGTTGCGATGATCAACCAATTAATAATTAATTAATTA 2882
QY 597 TAAAGTTGATCTTCAGTAAGTGGGCTTTCAAGAAATCAAGATTTGGCTCAGAAAT 656
DB 2883 TAGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2942
QY 657 TGACAATCTCATCA 671
DB 2943 TAAATTAATTAATTA 2957

RESULT 6
US-10-615-668-22

Sequence 22, Application US/10615668
Publication No. US20050276819A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides
FILE REFERENCE: CHIR0337
CURRENT APPLICATION NUMBER: US/10/615,668
PRIOR FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/410,835
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.2
SEQ ID NO 22
LENGTH: 45
TYPE: DNA
ORGANISM: Helicobacter pylori
US-10-615-668-22


```
Db      340 ATACGAGAGCTGTAAGAAAAATGATGATGATCAAAAAACAAGCAAGAAAGTGAAA 281
Qy      239 ATGCTAGAAAAAATCTGAATATATCAATCCGTAGAAATGCTGAATGGAACCTTAG 298
Db      280 GAATATATAATATTTTGGACATATAGGAATGAAATTTGATATGATGATTTAAACCAAA 221
Qy      299 TCGGTAATGGTTATCTCAAGCAGACCAACTCTTTCTAAAACTTTTCGAGACATCA 358
Db      220 TATGGAGTACTTACATTAAGATGACAGGCACAAAATTAACAAAATTTGTTGGAAGTTG 161
Qy      359 AGAAAGGTGATGACAAACTTGAATTTCAATTAACAATTAATGAGACTCAAAA 418
Db      160 TTGAATTTGATGAAAAATCATGCTCATCAAGACTGATATCTTAAACATATGATCTC 101
Qy      419 ACGAACCCATTATGCTAAGTTAATTAATAAGAA 453
Db      100 AGAATGTTTCAAAATTAATAAGATGAAAAAAGATA 66
```

RESULT 9

```
US-09-925-065A-127945/c
; Sequence 127945, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127945
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-127945
```

Query Match 6.2%; Score 42.2; DB 6; Length 533;

Best Local Similarity 54.2%; Pred. No. 2.5; Matches 83; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

```
Qy      305 ATGGGTTATCTCAAGCAGACCAACTCTTTCTAAAACTTTTCGAGACATCAAGAAAG 364
Db      260 ATTAGGTGTTAAAGGTAATGACATATTAAGTAAGTAATTTTGAACCTAAGAAACG 201
Qy      365 AGTTGATGCAAAACTTGAATTTCAATTAACAATTAATGACTCAAAAAGAAC 424
Db      200 TTAGAAATCATGATGAAATTTTTCACAGATAAAAAGAGTGATTCAAATCTCTC 141
Qy      425 CCATTATGCTAATGTTAATTAATAAGAAAGACG 457
Db      140 CAATTAATTCACAAATTAATAAGAAATTAAGAACTG 108
```

RESULT 10

```
US-09-925-065A-140562/c
; Sequence 140562, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
```

```
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140562
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-140562
```

Query Match 6.2%; Score 42.2; DB 6; Length 633;

Best Local Similarity 55.1%; Pred. No. 2.6; Matches 103; Conservative 0; Mismatches 83; Indels 1; Gaps 1;

```
Qy      302 GTAATGGTTATCTCAAGCAGACCAACTCTTTCTAAAACTTTTCGAGACATCAAG 361
Db      475 GCAAAAAGTTATCTTATGCAAAATGCAAAATGCAACAAAGACTTATCATTAAGACA 416
Qy      362 AAGAGTTGATGCAAAACTTGAATTTCAATTAACAATTAATGACTCAAAAAGC 421
Db      415 AATAGTTTACTTAAATTAATTAACAAGACAAATTAAC-ATTCTGTAATGATTAAG 357
Qy      422 AACCCATTATGCTAAGTTAATTAATAAGAAAGCAGGACGAGCTAGCTGAAGAC 481
Db      356 AATCAATTCATCGAAGAGTACATTAATTAACAATTAATGCACTTATCATTAAGAC 297
Qy      482 CCATTTA 488
Db      296 ATTAATA 290
```

RESULT 11

```
US-09-925-065A-575490
; Sequence 575490, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575490
; LENGTH: 524
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-575490
```

Query Match

6.0%; Score 41.2; DB 6; Length 524;

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rmpn

Page 6

Best Local Similarity 46.2%; Pred. No. 4.2; Mismatches 194; Indels 1; Gaps 1;
Matches 168; Conservative 1;

Qy 194 CAAAGAGCAATGGCCCAACACCACTCAAAAAAATGAACTGCAATGCGAATAAT 253
Db 36 CAAACAAACAAAAAGTACAAAATTTTGTAAATATGATTTAAGAAACCAAAAAA 95
Qy 254 CTGAATATATCATCGCTTAAGATGCTGCAATGACCCCTAGTCGTAAAGGCTTAT 313
Db 96 AAAAATCAATATCTGAAATATATTAATTTAAGCAATGACGCACTTACAAA 155
Qy 314 CTCAGAGAGAGCCCAACCTCTTTTAAATTTTGGACATCAAGAAAGAGTTATG 373
Db 156 TCTATGAGACATAGAAATATGTAATTAATTAATTAATTAATTAATTAATTA 214
Qy 374 CAAAGCTTGAATTTTCAATACATTAATTAATTAATTAATTAATTAATTAAT 433
Db 215 TATATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 274
Qy 434 CTAAGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 493
Db 275 CTACTGATCTACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 334
Qy 494 AAGTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 553
Db 335 CAGCTCAACAGAGCTTATCAACAAAAATCAATATGATCACTAAGAAATTA 394
Qy 554 GTGT 557
Db 395 GTTT 398

RESULT 12

US-10-240-708-1/c
Sequence 1, Application US/10240708
Publication No. US20050282157A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIERENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/240,708
PRIOR FILING DATE: 2002-10-03
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-10
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 1
LENGTH: 10467
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-1

Query Match
Best Local Similarity 5.9%; Score 40.6; DB 8; Length 10467;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 345 CTTTGGACATCAAGAAAGTTGATGATGAACTTGAATTTCAATTAACAATAACA 404
Db 10178 CTTTAAACAAATTTAAAAATTAATTAATTTCTAATTTCAACAAATTAATCC 10119
Qy 405 TAATGAGCTCAAAACCAACCAATTTATGTAATTAATTAATTAATTAATTAAT 464

Db 10118 CTTTAAATCBAACACCAATATATATTTATTTTAAAAACAAAAATTAACCTA 10059
Qy 465 AGTACGCTTGAAGACCCATTTTACGCTCAAGTTGCTAAAGAAATGCAAAATGTA 524
Db 10058 AATTAACAAACAAATCAATCAATTAATTAATTAATTAATTAATTAATTAAT 9999
Qy 525 CCGAGCTCATC 535
Db 9998 ACGAAATTAAC 9998

RESULT 13

US-09-925-065A-410917
Sequence 410917, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 410917
LENGTH: 598
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-410917

Query Match
Best Local Similarity 49.8%; Pred. No. 9.1; Mismatches 123; Indels 1; Gaps 1;
Matches 124; Conservative 1;

Qy 195 AAAGAGCAATGGCCCAACACCACTCAAAAAAATGAAAGTCTCAATGTAAGAAAAATC 254
Db 114 AATTAACAAACAAAAAGTACAAAATTTTGTAAATATGATTTAAGAAACCAAAAAA 173
Qy 255 TGAATATATCAATCCGTTAAGATGCTGTAATGAAACCTAGTCGTAATGCTTATC 314
Db 174 TGAACAAATATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 233
Qy 315 TCAAGCAAGAGCCCAACCTCTTTTAAATTTTGGACATCAAGAAAGAGTTAATG 374
Db 234 TTAATAGGTAGTACTTTTGTGAARAAAGACTTCCAAAGTACATTAATTAATG 292
Qy 375 AAACTTGAATTTCAATTAACATTAATTAATTAATTAATTAATTAATTAATTA 434
Db 293 AAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 352
Qy 435 TAAAGTTAA 443
Db 353 CAAAAATTA 361

RESULT 14

US-09-925-065A-474252/c
Sequence 474252, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single

;; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.135
;; CURRENT APPLICATION NUMBER: US/09/925,065A
;; CURRENT FILING DATE: 2001-08-08
;; PRIOR APPLICATION NUMBER: US 60/243,096
;; PRIOR FILING DATE: 2000-10-24
;; PRIOR APPLICATION NUMBER: US 60/252,147
;; PRIOR FILING DATE: 2000-11-20
;; PRIOR APPLICATION NUMBER: US 60/250,092
;; PRIOR FILING DATE: 2000-11-30
;; PRIOR APPLICATION NUMBER: US 60/261,766
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/289,846
;; PRIOR FILING DATE: 2001-05-09
;; NUMBER OF SEQ ID NOS: 957086
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 474252
;; LENGTH: 670
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-925-065A-474252

Query Match 5.8%; Score 39.8; DB 6; Length 670;
Best Local Similarity 54.4%; Pred. No. 9.5;
Matches 80; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 375 AAACCTGAAATTTCAATACATATATGACCTGAAAGCAACCCATTATGC 434
DB 584 AAAGCAAGAAATTAAGAAATGCGATTTGAAAAAGACCAAGAAATCTCCATATATGA 525
QY 435 TAAAGTTAATTAAGAAAGCAGGCGACGCTAGCTTGAAGAACCCATTACGCTCA 494
DB 524 AAAATATATTAATTTGAATCCAACTTAACAGATACGATTAAAAATTAACATTATAGAAAT 465
QY 495 AGTTGCTAAAGGTTAATGCAAAAT 521
DB 464 AGCTGATTAAGAAAGACATTTTAAAAAT 438

RESULT 15
US-09-925-065A-474253/c
;; Sequence 474253, Application US/09925065A
;; Publication No. US20040181048A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single
;; FILE REFERENCE: 108827.135
;; Nucleotide Polymorphisms in the Human Genome
;; CURRENT APPLICATION NUMBER: US/09/925,065A
;; CURRENT FILING DATE: 2001-08-08
;; PRIOR APPLICATION NUMBER: US 60/243,096
;; PRIOR FILING DATE: 2000-10-24
;; PRIOR APPLICATION NUMBER: US 60/252,147
;; PRIOR FILING DATE: 2000-11-20
;; PRIOR APPLICATION NUMBER: US 60/250,092
;; PRIOR FILING DATE: 2000-11-30
;; PRIOR APPLICATION NUMBER: US 60/261,766
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/289,846
;; PRIOR FILING DATE: 2001-05-09
;; NUMBER OF SEQ ID NOS: 957086
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 474253
;; LENGTH: 670
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-925-065A-474253

Query Match 5.8%; Score 39.8; DB 6; Length 670;
Best Local Similarity 54.4%; Pred. No. 9.5;
Matches 80; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 375 AAACCTGAAATTTCAATACATATATGACCTGAAAGCAACCCATTATGC 434

DB 584 AAAGCAAGAAATTAAGAAATGCGATTTGAAAAAGACCAAGAAATCTCCATATATGA 525
QY 435 TAAAGTTAATTAAGAAAGCAGGCGACGCTAGCTTGAAGAACCCATTACGCTCA 494
DB 524 AAAATATATTAATTTGAATCCAACTTAACAGATACGATTAAAAATTAACATTATAGAAAT 465
QY 495 AGTTGCTAAAGGTTAATGCAAAAT 521
DB 464 AGCTGATTAAGAAAGACATTTTAAAAAT 438

Search completed: March 6, 2006, 22:31:20
Job time : 352 secs

This Page Blank (uspto)

Tue Mar 7 12:58:26 2006

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 22:06:00 ; Search time 782 Seconds

(without alignments)
7243.637 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685

Sequence: 1 aaaaatgcgaataaaga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NIC

Gapop 10.0, Gapept 1.0

Searched: 9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq.*
5: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	100.0	4149	3	US-09-402-100-3 Sequence 3, Appl1
2	685	100.0	5925	3	US-09-921-157-4 Sequence 4, Appl1
3	477.2	69.7	3504	7	US-10-335-977-3871 Sequence 3871, Ap
4	477.2	69.7	3534	7	US-10-335-977-3872 Sequence 3872, Ap
5	477.2	69.7	3549	7	US-10-335-977-3873 Sequence 3873, Ap
6	50.6	7.4	3673778	6	US-10-312-841-1 Sequence 1, Appl1
7	48.2	7.0	571	4	US-09-925-065A-20908 Sequence 20908, A
8	48	7.0	12138	6	US-10-311-455-1916 Sequence 1916, Ap
9	48	7.0	12138	6	US-10-240-453-210 Sequence 210, Appl1
10	47.8	7.0	3673778	6	US-10-312-841-2 Sequence 1931, Ap
11	47	6.9	14006	6	US-10-311-455-1993 Sequence 1993, Ap
12	46.8	6.8	5542	6	US-10-311-455-1993 Sequence 71, Appl1
13	46.8	6.8	9964	6	US-10-311-455-71 Sequence 38710, A
14	46.4	6.8	921	8	US-10-425-115-38710 Sequence 132, Appl1
15	46.4	6.8	5127	5	US-10-239-676-132 Sequence 401, Appl1
16	46	6.7	7969	7	US-10-221-714A-401 Sequence 67, Appl1
17	46	6.7	7969	7	US-10-433-793-67 Sequence 1988, Ap
18	45.6	6.7	6182	6	US-10-311-455-1988 Sequence 78, Appl1
19	45.4	6.6	14362	7	US-10-333-793-78 Sequence 222, Appl1
20	45.4	6.6	14708	5	US-10-239-676-222 Sequence 2216, Ap
21	45.4	6.6	14708	5	US-10-311-455-324 Sequence 324, Appl1
22	45.4	6.6	14708	6	US-10-240-453-324 Sequence 500, Appl1
23	45.4	6.6	14708	7	US-10-221-714A-500 Sequence 500, Appl1

C 24	45.4	6.6	14708	7	US-10-240-589C-134	Sequence 134, Appl1
C 25	45.2	6.6	8222	9	US-10-486-319A-43	Sequence 43, Appl1
C 26	45.2	6.6	8222	9	US-10-486-319A-65	Sequence 65, Appl1
C 27	45	6.6	5814	6	US-10-311-455-1534	Sequence 1534, Ap
C 28	45	6.6	9666	6	US-10-311-455-1058	Sequence 1058, Ap
C 29	45	6.6	9666	6	US-10-240-485-84	Sequence 84, Appl1
C 30	45	6.6	11812	5	US-10-239-676-210	Sequence 210, Appl1
C 31	45	6.6	11812	6	US-10-311-455-2092	Sequence 2092, Ap
C 32	45	6.6	11812	6	US-10-240-453-306	Sequence 306, Appl1
C 33	45	6.6	11812	7	US-10-221-714A-468	Sequence 468, Appl1
C 34	45	6.6	19787	6	US-10-311-455-1423	Sequence 1423, Ap
C 35	44.8	6.5	1332	2	US-08-325-278-5	Sequence 5, Appl1
C 36	44.8	6.5	1662	8	US-10-474-792-671	Sequence 671, Appl1
C 37	44.8	6.5	2543	5	US-10-027-632-102481	Sequence 102481, Appl1
C 38	44.8	6.5	2543	5	US-10-027-632-102482	Sequence 102482, Appl1
C 39	44.8	6.5	2543	6	US-10-027-632-102481	Sequence 102481, Appl1
C 40	44.8	6.5	2543	6	US-10-027-632-102482	Sequence 102482, Appl1
C 41	44.8	6.5	3885	8	US-10-377-636-3	Sequence 3, Appl1
C 42	44.6	6.5	113515	6	US-10-311-455-2147	Sequence 2147, Ap
C 43	44.2	6.5	6171	6	US-10-311-455-761	Sequence 761, Appl1
C 44	44.2	6.5	7369	8	US-10-473-126-277	Sequence 277, Appl1
C 45	44.2	6.5	7369	9	US-10-486-319A-56	Sequence 56, Appl1

ALIGNMENTS

```

RESULT 1
US-09-402-100-3
; Sequence 3, Application US/09402100
; Patent No. US20010019834A1
; GENERAL INFORMATION:
; APPLICANT: Daewoong Pharmaceutical Co, LTD
; APPLICANT: Kim, Byung-O
; APPLICANT: Shin, Sung-Seup
; APPLICANT: Yu, Young-Hyo
; APPLICANT: Park, Myung-Hwan
; APPLICANT: Choi, Deok-Oon
; APPLICANT: Jung, Hyung-Jin
; TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Heli.
; FILE REFERENCE: 0136/06140
; CURRENT APPLICATION NUMBER: US/09/402,100
; CURRENT FILING DATE: 1999-09-27
; EARLIER APPLICATION NUMBER: KR 97-11950
; EARLIER FILING DATE: 1997-03-31
; EARLIER APPLICATION NUMBER: KR 97-11951
; EARLIER FILING DATE: 1997-03-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1..4149)
; OTHER INFORMATION: Recombinant DNA
US-09-402-100-3
Query Match 100.0%; Score 685; DB 3; Length 4149;
Best Local Similarity 100.0%; Pred. No. 3.1e-153;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAGCAAAAATGATTCGCAAGTAACGCAAGCAAAAAGGACCTGAAT 60
DB AAAAAAGCAAAAATGATTCGCAAGTAACGCAAGCAAAAAGGACCTGAAT 2307
QY 61 TCCGTTAAAGATGATCATCATCAAAAGGTAACGATTAAGTATCTCAATCA 120
DB TCCGTTAAAGATGATCATCATCAAAAGGTAACGATTAAGTATCTCAATCA 2367
QY 121 GCGGATCATGCGGCTTAAGCAAGCGGATTCGATGAGGTAGCAAGCGGTA 180

```

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rnpsm

Page 2

Db 2368 GCGGTATCAGTGGCTTAAGCAAGCGGTGATTTCAGTACGGTAAAGCAGCTTAAAGCCGAT 2427
QY 181 CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 240
Db 2428 CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 2487
QY 241 GCTGAAAAAAATCTGAAATTAATCAATCCCTTAAGATGGTGAATGAACCTTATGTC 300
Db 2488 GCTGAAAAAAATCTGAAATTAATCAATCCCTTAAGATGGTGAATGAACCTTATGTC 300
QY 301 GGTATAGGCTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
Db 2548 GGTATAGGCTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
QY 361 AAGAGTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAACATTAACATTA 420
Db 2608 AAGAGTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAACATTAACATTA 420
QY 421 GAACGATTTATCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 2667
Db 2668 GAACGATTTATCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 2727
QY 481 CCCATTTACGCTCAAGTGTCTAAAGAGTAAATGCAAAATTTGACCGACTCAATCAATA 540
Db 2728 CCCATTTACGCTCAAGTGTCTAAAGAGTAAATGCAAAATTTGACCGACTCAATCAATA 540
QY 541 GCAAGTGTGTGAGT 600
Db 2788 GCAAGTGTGTGAGT 600
QY 601 GTTGATGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
Db 2848 GTTGATGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
Db 2908 AATCTCAATCAAGCGGTATCAAGAG 2932

RESULT 2

US-09-921-157-4
Sequence 4, Application US/09921157
Publication No. US2004004835A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonio
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For
CURRENT INVENTION: Vaccines And Diagnostics
CURRENT APPLICATION NUMBER: US/09/921,157
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/466,662
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 08/360,934
PRIOR FILING DATE: 1993-07-26
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1993-06-06
PRIOR APPLICATION NUMBER: PCT/EP93/00472
PRIOR FILING DATE: 1993-03-02
PRIOR APPLICATION NUMBER: PCT/EP93/00158
PRIOR FILING DATE: 1993-01-25
PRIOR APPLICATION NUMBER: PCT/EP93/00158
PRIOR FILING DATE: 1992-03-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 4
LENGTH: 5925
TYPE: DNA

ORGANISM: Helicobacter pylori

US-09-921-157-4

Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 3,6e-153;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAATGCAAAATTAAGATTTCAAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 60
Db 2782 AAAAAATGCAAAATTAAGATTTCAAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2841
QY 61 TCCGTTAAGATGATCATCATCAAAAGGTAAAGTTGATTAATCATCAATCA 120
Db 2842 TCCGTTAAGATGATCATCATCAAAAGGTAAAGTTGATTAATCATCAATCA 120
QY 121 GCGGTATCAAGTGTAAAGCAAGGATGATTTCAAGAGTGAAGCAAGGATGATTTCAAGAGTGAAG 180
Db 2902 GCGGTATCAAGTGTAAAGCAAGGATGATTTCAAGAGTGAAGCAAGGATGATTTCAAGAGTGAAG 180
QY 181 CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 240
Db 2962 CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 240
QY 241 GCTAAGAAAAATCTGAATATATCATCCGTTAAGATGATGATGATGATGATGATGATGATGATGAT 300
Db 3022 GCTAAGAAAAATCTGAATATATCATCCGTTAAGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GGTATAGGTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
Db 3082 GGTATAGGTTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
QY 361 AAGAGTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAACATTAACATTAACATTA 420
Db 3142 AAGAGTGAATGCAAACTTGAAATTTCAATTAACATTAACATTAACATTAACATTAACATTAACATTA 420
QY 421 GAACGATTTATCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 540
Db 3202 GAACGATTTATCTCAAAAGAGCAATTTGGCCCAACACCTCAAAAAATGAAGTCTCAAT 540
QY 481 CCCATTTACGCTCAAGTGTCTAAAGAGTAAATGCAAAATTTGACCGACTCAATCAATA 600
Db 3262 CCCATTTACGCTCAAGTGTCTAAAGAGTAAATGCAAAATTTGACCGACTCAATCAATA 600
QY 541 GCAAGTGTGTGAGT 660
Db 3322 GCAAGTGTGTGAGT 660
QY 601 GTTGATGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
Db 3382 GTTGATGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
Db 3442 AATCTCAATCAAGCGGTATCAAGAG 3466

RESULT 3

US-10-335-977-3871
Sequence 3871, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBLACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rnpbm

Page 4

```
Db 2353 TCATTAAGATGTGATCATCATCAAAAGATACGATTAAGTTGATATCTCATCA 2412
Qy 121 GCGGTATCAATGCTGAAGCAAGCGGTGATTTCAAGTGGGTAGACCAAGCTTACCGAT 180
Db 2413 GCGGTATCAATGCTGAAGCAAGCGGTGATTTCAAGTGGGTAGACCAAGCGGTACCGAT 2472
Qy 181 CTCAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTTCAAAAAATGAAATCTCAAT 240
Db 2473 CTCAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTTCAAAAAATGAAATCTCAAT 2532
Qy 241 GCTAGAAAAATTCGAAATATTCATTCGTTTAAAGATGTGATGAAACCTTAGTC 300
Db 2533 ACTGAAAAATTTCTGCACTATACCATTCGTTAAAGATGTGATGAAACCTTAGTC 2592
Qy 301 GGTAAAGGTTATCTCAAGCAAGCAACCTTTCTTAAATCTTTGCAATCAAG 360
Db 2593 GGTAAAGGTTATCTCAAGCAAGCAACCTTTCTTAAATCTTTGCAATCAAG 2652
Qy 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAGTCAATCAAAAC 420
Db 2653 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAGTCAATCAAAAC 2712
Qy 421 GAACCATTTATGCTAAAGTTATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
Db 2713 AGCAC-----MGA 2721
Qy 481 CCATTTACGCTCAAGTTGCTTAAAGATTAATGCAAAATTTGACGACTCAATCAATA 540
Db 2722 CCATTTATATCTCAAGTTGCTTAAAGATTAATGCAAAATTTGACGACTCAATCAATA 2781
Qy 541 GCAAGTGTGTTGGGTGTGTAGGCGCAAGCGGCGCTTCCCTTGAAGAGCATGATAA 600
Db 2782 GCAAGTGTGTTGGGTGTGTAGGCGCAAGCGGCGCTTCCCTTGAAGAGCATGATAA 2841
Qy 601 GTTGATGATCTCAGTAAAGTAAAGGCTTTCAGAGATCAAGAA 642
Db 2842 GTTGATGATCTCAGTAAAGTAAAGGCTTTCAGAGATCAAGAA 2893
```

```
RESULT 5
US-10-335-977-3873
; Sequence 3873, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHYE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-2875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335, 977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993, 002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandagouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
```

```
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3873:
SEQUENCE CHARACTERISTICS:
LENGTH: 3549 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...3549
SEQUENCE DESCRIPTION: SEQ ID NO: 3873:
US-10-335-977-3873
```

```
Query Match 69.7%; Score 477.2; DB 7; Length 3549;
Best Local Similarity 86.9%; Pred. No. 1,5e-103;
Matches 558; Conservative 0; Mismatches 33; Indels 51; Gaps 1;
```

```
Qy 1 AAAAAATGCAAAATTAAGGATTTGACAGTTACGCAAGCAAAAGCAAGCTTGAAGAT 60
Db 2308 AAAAAATGCAAAATTAAGGATTTGACAGTTACGCAAGCAAAAGCAAGCTTGAAGAT 2367
Qy 61 TCGTTAAAGATGTGATCATCATCAAAAGTTAACGATTAAGTTGATATCTCAATCA 120
Db 2368 TCGATTAAGATGTGATCATCATCAAAAGTTAACGATTAAGTTGATATCTCAATCA 2427
Qy 121 GGGGTATCATGCTTAAAGCAAGGATTTTCAAGTGGTAAAGCAAGCTTACGCGAT 180
Db 2428 GGGGTATCATGCTTAAAGCAAGGATTTTCAAGTGGTAAAGCAAGCTTACGCGAT 2487
Qy 181 CTCAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTTCAAAAAATGAAAGTCTCAAT 240
Db 2488 CTCAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTTCAAAAAATGAAAGTCTCAAT 2547
Qy 241 GCTGAAAAAATCTGAAATATATTCATTCCTTAAAGATGTGATGATGACCTTAGTC 300
Db 2548 ACTGAAAAAATCTGCACTATACCATTCCTTAAAGATGTGATGATGACCTTAGTC 2607
Qy 301 GGTAAAGGTTATCTCAAGCAAGCAACCTTTCTTAAATCTTTGCAATCAAG 360
Db 2608 GGTAAAGGTTATCTCAAGCAAGCAACCTTTCTTAAATCTTTGCAATCAAG 2667
Qy 361 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAGTCAATCAAAAC 420
Db 2668 AAAGAGTTGAATGCAAACTTGAAATTTCAATTAACATTAATGAGTCAATCAAAAC 2727
Qy 421 GAACCATTTATGCTAAAGTTATTAAGAAAGCAAGGCAAGCAAGCTTGAAGAA 480
Db 2728 AGCAC-----MGA 2736
Qy 481 CCATTTACGCTCAAGTTGCTTAAAGATTAATGCAAAATTTGACGACTCAATCAATA 540
Db 2737 CCATTTATATCTCAAGTTGCTTAAAGATTAATGCAAAATTTGACGACTCAATCAATA 2796
Qy 541 GCAAGTGTGTTGGGTGTGTAGGCGCAAGCGGCGCTTCCCTTGAAGAGCATGATAA 600
Db 2797 GCAAGTGTGTTGGGTGTGTAGGCGCAAGCGGCGCTTCCCTTGAAGAGCATGATAA 2856
Qy 601 GTTGATGATCTCAGTAAAGTAAAGGCTTTCAGAGATCAAGAA 642
Db 2857 GTTGATGATCTCAGTAAAGTAAAGGCTTTCAGAGATCAAGAA 2898
```

```
RESULT 6
US-10-312-841-1/C
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Bpigenomics AG
```


; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
 ; FILE REFERENCE: E01/1208/MO
 ; CURRENT APPLICATION NUMBER: US/10/312,841
 ; CURRENT FILING DATE: 2002-12-30
 ; NUMBER OF SEQ ID NOS: 2
 ; SEQ ID NO 1
 ; LENGTH: 3673778
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 ; NAME/KEY: unsure
 ; LOCATION: (3294164)
 ; US-10-312-841-1

Query Match 7.4%; Score 50.6; DB 6; Length 3673778;
 Best Local Similarity 47.4%; Pred. No. 3.3;
 Matches 152; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 220 CAAAAAATGAAAGTCTCAATGCTAGAAAAAATCGAATATATCAATCCGTTAAGAT 279
 DB 2724190 CACTTAAAAAACAATCAATTTATTAATAAAAAAATAATTCACCCCACTTTAAC 2724131
 QY 280 GGTGTAATGGAACCTAGTCGGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCT 339
 DB 2724130 TCATATTTTAAAAAACTCTACTTAAAAAATACATTCATCACTTAACCAAAATTAAT 2724071
 QY 340 AAAAATTTTGGACATCAAGAAAGTTGATGCAAACTTGAATTTCAATACAT 399
 DB 2724070 AAAAAACATATATAAATCTAAAAAAGAAATTCAAAAAACAATAATTAACAAA 2724011
 QY 400 AACAAATATGAGTCAAAAAAGAACCATTTATGCTAAAGTATATTAAGAAAGCAAGG 459
 DB 2724010 AAAAAATAAAAACCTTAATTAATAATTAATACAAAAAACAATAAATAAATAAATA 2723951
 QY 460 CAAGCAGTACGCTTGAAGAACCATTTACGCTCAAGTGTCTAAAAAGTAATGAAAA 519
 DB 2723950 CTAACTATTAACCTTCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2723891
 QY 520 ATTGACCACTCAATCAATA 540
 DB 2723890 ATCAAAAACCTCATATATTA 2723870

RESULT 7

; US-09-925-065A-20908
 ; Sequence 20908, Application US/09925065A
 ; Publication No. US20050228172A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single
 ; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
 ; CURRENT APPLICATION NUMBER: US/09/925,065A
 ; CURRENT FILING DATE: 2001-08-08
 ; PRIOR APPLICATION NUMBER: US 60/243,096
 ; PRIOR FILING DATE: 2000-10-24
 ; PRIOR APPLICATION NUMBER: US 60/252,147
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/250,092
 ; PRIOR FILING DATE: 2000-11-30
 ; PRIOR APPLICATION NUMBER: US 60/261,766
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/289,846
 ; PRIOR FILING DATE: 2001-05-09
 ; NUMBER OF SEQ ID NOS: 957086
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 20908
 ; LENGTH: 571
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-925-065A-20908

Query Match 7.0%; Score 48.2; DB 4; Length 571;
 Best Local Similarity 50.7%; Pred. No. 0.27;
 Matches 116; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 221 AAAAAAATGAAAGTCTCAATGCTAGAAAAAATCGAATATATCAATCCGTTAAGAT 280
 DB 24 MAGAAAATCACTGTAATACCCGATAGAAAAGGGCAAAAGTCTTCAGCATTTACAAATG 83
 QY 281 GTGTGAATGGAACCTAGTCGGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCT 340
 DB 84 GTATGTTTGTAAATGAATGATTTGATCCATTAATTTGATCCATGCGCCCAATTAATATG 143
 QY 341 AAAAATTTTGGACATCAAGAAAGTTGATGCAAACTTGAATTTCAATACAT 400
 DB 144 AAAAGATATGATCACTTCAATCAGCCATTAAGAAAAATTAAGATTAATAAATGCAATGAGCTAA 203
 QY 401 ACATATATGAGCTCAAAAAGCAACCATTTATGCTAAAGTTAATAAAA 449
 DB 204 AACGATATGAAAAAAGAAATTAATCAATGTTTAAGAAAGATGTGAAAAA 252

RESULT 8

; US-10-311-455-1916/c
 ; Sequence 1916, Application US/10311455
 ; Publication No. US20030143606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De-
 ; FILE REFERENCE: 5013.1014
 ; CURRENT APPLICATION NUMBER: US/10/311,455
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537
 ; PRIOR FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 2424
 ; SEQ ID NO 1916
 ; LENGTH: 12138
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 ; US-10-311-455-1916

Query Match 7.0%; Score 48; DB 6; Length 12138;
 Best Local Similarity 47.5%; Pred. No. 1.1;
 Matches 174; Conservative 0; Mismatches 190; Indels 2; Gaps 1;

QY 161 TAGAGCAAGCTTACCGCATCTCAAAAATTTCTCAAGAGCAATGCGCCCAAGAGCTC 220
 DB 9844 TATATAAAATCACTAATATTAATAAACTCTTCTTAATTAATAAATTAATCAAAACAATCT 9785
 QY 221 AAAAAAATGAAAGTCTCAATGCTAGAAAAAATCGAATATATCAATCCGTTAAGAT 280
 DB 9784 TCAAAAAATTCATCTAATCACTCAAAATTAATAAATCACTTATTAATTAATTAATTAAT 9725
 QY 281 GTGTGAATGGAACCTAGTCGGTAATGSGTTATCTCAAGCAGAGCCCACTCTTCT 340
 DB 9724 CCCCTCTTAATTAATAATTTCTTAATCAATTTTATTAATAAATTAATCAAAACAATTTCA 9665
 QY 341 AAAAATTTTGGACATCAAGAAAGTTGATGCAAACTTGAATTTCAATACAT 400
 DB 9664 AAAAATTTTCAATTTCTAAAAATCAAAAAATTAATAAATTAATTAATTAATTAATTAAT 9605
 QY 401 ACAAT--AATGACATCAAAAAGCAACCATTTATGCTTAATTAATTAATAAAGCAAG 458
 DB 9604 AAAATCAACAACAAACAAACATTAACAAAAAATTAATTAATAAATAAATAAATAATCTAA 9545

QY 459 GCAAGCAGCTAGCCTTGAAGACCCATTAGCTCAAGTTGCTAAAAAGTAATGCANA 518
DB 9544 AAAACCACTATCAAAAAAABAAAAAACCATTATATATACCAAAAAAATCTTAA 9485
QY 519 AATTGA 524
DB 9484 TTTTAA 9479

RESULT 9
US-10-240-453-210/c
Sequence 210, Application US/10240453
Publication No. US20030146326A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
FILE REFERENCE: 5013.1009
CURRENT APPLICATION NUMBER: US/10/240.453
PRIOR FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 210
LENGTH: 12138
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-210

Query Match
Best Local Similarity 47.5%; Pred. No. 1.1;
Matches 174; Conservative 0; Mismatches 190; Indels 2; Gaps 1;

QY 161 TAGAGCAGGTTAGCCGATCTCAAAAATTTCTCAAGAGCAATGGCCCAACAGCTC 220
DB 9844 TATTAATACACTAATATTTAAATCTCTTATTAATAAAAAATATCAAAACATTA 9785
QY 221 AAAAAATGAAGTCTCAATGCTGAAAAAATCTGAATATATCAATCCGTTAAGATG 280
DB 9784 TCAAAAAATTCATCTAATCAATATTAATAATACCTTATTAATTAATTAATTA 9725
QY 281 GTGTGAATGAACCTTGTGCTAATGGTTATCTCAAGCAGAGCCCAACTTTTCA 340
DB 9724 CCCCTTATTAATTAATTTCTTAATCAATTTATTAATAAATTAACAAAACATTTCA 9665
QY 341 AAAATTTTCGACATCAAGAAAGTTGCAAAACCTTGAAATTTCAATTAACATA 400
DB 9664 AAAATTAATTTCAATTTTAAAAATTAATAAATTAATAATTAATTAATTA 9605
QY 401 ACAT--AATGACTCAAAAGCAACCATTTTGTGAAGTAAATTAATAAAGAAAGCAAG 458
DB 9604 AAAATTAACAAACAAACATTAACAAAATTAATAAATAAATAAATAAATAAATA 9545
QY 459 GCAAGCAGCTAGCCTTGAAGAACCCATTTAGCTCAAGTTGCTAAAAAGTAATGCANA 518
DB 9544 AAAACCACTATCAAAAAAABAAAAAACCATTATATATACCAAAAAAATCTTAA 9485
QY 519 AATTGA 524

DB 9484 TTTTAA 9479

RESULT 10
US-10-312-841-2/c
Sequence 2, Application US/10312841
Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomix AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MR
FILE REFERENCE: E01/1208/WO
CURRENT APPLICATION NUMBER: US/10/312.841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 2
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (379615)
US-10-312-841-2

Query Match
Best Local Similarity 46.5%; Pred. No. 15;
Matches 154; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 183 CAAAAATTTCTCAAGAGCAATGGCCCAACCTCAAAAAATGAAGTCTCAATGC 242
DB 3055754 CAAAAATTAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA 3055695
QY 243 TAGAAAAATCTGAATATATATCAATCCGTTAAGATGGTGAATGAAACCCCTAGTCGG 302
DB 3055694 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055635
QY 303 TAATGGTATCTCAAGAGCAAGCCCAACTTTTCTAAAACTTTTGGACATCAAGAA 362
DB 3055634 ATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055575
QY 363 AGATTAATGAACAACTTGAAATTTCAATTAACATTAATTAATTAATTAATTAATTA 422
DB 3055574 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055515
QY 423 ACCATTTTGTGAAGTAAATTAATAAAGAAAGCAAGGCAAGCTAGCTGAAGAAC 482
DB 3055514 AAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055455
QY 483 CATTTAGCTCAAGTCTTAATAAAGTAAT 513
DB 3055454 AATTTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3055424

RESULT 11
US-10-311-455-1931/c
Sequence 1931, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting
TITLE OF INVENTION: Cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1931
LENGTH: 14006
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: 8289, 8310, 8313
OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931

Query Match
Best Local Similarity 44.7%; Score 47; DB 6; Length 14006;
Pred. No. 2.1; Mismatches 275; Indels 1; Gaps 1;
Matches 223; Conservative 0;

42 AAAAAAGCAGCTTGAATAATTCGTTAAGATGATCATCAATCAAAAGGTACGGATA 101
6018 AAAAAAAAAAATTAACAACTTAACAAAAATCAAACTAAATTAATTAATTAATTA 5959
102 AGTATATCTCAATCAAGCGGTATCAAGTGGCTAAGCAACGGGTATTCAGTAGGT 161
5958 ATTATATTAATAAATCAAACTAATCAAAATTAATTAATTAATTAATTAATTAAT 5900
162 AAGAGAGGCTTACCGATCTCAAAATTTCTCAAGAGCAATTTGCCCAACAGCTCA 221
5899 ATTATCTGACTTCATCAATCAAAATTAATTAATTAATTAATTAATTAATTAATTA 5840
222 AAAAAATGAAGCTCAATGCTGAATAAATCTGAATATATCAATCGTTAAGATGG 281
5839 AATATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5780
282 TGTGAATGACCTTGAAGTGGTATCTCAAGCAAGCAACTCTTTCTAA 341
5779 CCAAAATACATATATACCAAAATCTTAATTAATTAATTAATTAATTAATTAATTA 5720
342 AACTTTTGGACATCAAGAAAGTTGAATGCAAACTTGAATTTCAATTAACATTA 401
5719 CAATTTCTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5660
402 CAATATGAGCTCAAAAGCAACCCATTAATGCTAAGTTAATTAATAAAGAGCGGCA 461
5659 CAACACTATTTCTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5600
462 AGCAGTACCTTGAAGAACCTTCAAGTTCAGTTGCTAATAAAGTAAATGCAAAAT 521
5599 AATTAACCTTCTAATTAACATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5540
522 TGACGAGCTCAATCAATA 540
5539 AAAAAATTCACATCTAATA 5521

RESULT 12
US-10-311-455-1993/c
Sequence 1993, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIRENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1993
LENGTH: 5542
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1993

Query Match
Best Local Similarity 48.2%; Score 46.8; DB 6; Length 5542;
Pred. No. 1.6; Mismatches 142; Indels 0; Gaps 0;
Matches 132; Conservative 0;

180 TCTCAAAATTTCTGAAGAGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCA 239
4485 TCACATATATTCAAAAATCTAAATCAATTAATTAATTAATTAATTAATTAATTAAT 4426
240 TGTGAAAAAATTCGAATATATCAATCGTTAAGATGATGATGAATGCAACCTTGT 299
4425 TCATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4366
300 CGGTATGAGTATCTCAAGAGCAAGCAACTCTTTCAAAATCTTTGGACATCA 359
4365 CAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4306
360 GAAAGAGTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTAATTA 419
4305 AAGTTTATACATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4246
420 CGAACCAATTAATGCTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 453
4245 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4212

RESULT 13
US-10-311-455-71/c
Sequence 71, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIRENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 71
LENGTH: 9964
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

Query Match
Best Local Similarity 45.4%; Score 46.8; DB 6; Length 9964;
Pred. No. 2; Mismatches 292; Indels 5; Gaps 2;
Matches 247; Conservative 0;

1 AAAAAAGCAGCTTGAATAATTCGTTAAGATGATCATCAATCAAAAGGTACGGATA 60
9281 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9222
61 TCCGTAAGATGATCATCAATCAAAAGGTAAAGGATTAAGTGAATTAATCAATCA 120

Db 9221 AAAAAAAAAAATCAAAAAAAAAAAAAAAAAAATTAACACACAACTCATATAA 9162
Qy 121 GCGGTATCGTGGCTAAAGCAAGGGTATTTCTGATGGTAGACAAAGCTTAGCCGAT 180
Db 9161 TATTTATTAATTTTAAATTA-----ATTAAATTAATTAACAAATCTTAATTAATTA 9106
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAACTCAAAAAATGAAGTCTCAAT 240
Db 9105 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9046
Qy 241 GCTAGAAAAAATCTGAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 9046
Db 9045 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8986
Qy 301 GGTATGGTTATCTCAAGCAAGCCCAACTCTTTCTAAAACTTTTCGACATCAAG 360
Db 8985 AAAAAAAAAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8926
Qy 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTAAT 420
Db 8925 ATCGAAAAAATAA-AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8867
Qy 421 GAACCATTTATCTAAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
Db 8866 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8807
Qy 481 CCCATTTAGCTCAAGTGTCTAAAGTTAATTAATTAATTAATTAATTAATTAAT 540
Db 8806 TAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
Qy 541 GCAG 544
Db 8746 TAAA 8743

RESULT 14
US-10-425-115-38710/c
Sequence 38710, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 38710
LENGTH: 921
TYPE: DNA
ORGANISM: Zee mays
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(921)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match 6.8%; Score 46.4; DB 8; Length 921;
Best Local Similarity 42.8%; Fred. No. 0.9;
Matches 233; Conservative 0; Mismatches 311; Indels 0; Gaps 0;
Qy 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGTACGCAAGCAAAAGCACTTGAAT 60
Db 851 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 792
Qy 61 TCCGTAAAGATGTATCATCAATCAAAAGTAACGATTAATGAATATCTATCA 120
Db 791 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 732

Qy 121 GCGGTATCGTGGCTAAAGCAAGGGTATTTCTGATGGTAGACAAAGCTTAGCCGAT 180
Db 731 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 672
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAACAACTCAAAAAATGAAGTCTCAAT 240
Db 671 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 612
Qy 241 GCTAGAAAAAATCTGAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 300
Db 611 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
Qy 301 GGTATGGTTATCTCAAGCAAGCCCAACTCTTTCTAAAACTTTTCGACATCAAG 360
Db 551 AAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 492
Qy 361 AAAGAGTTGAATGCAAACTTGAATTTCAATTAATTAATTAATTAATTAATTAAT 420
Db 491 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 432
Qy 421 GAACCATTTATCTAAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
Db 431 AAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 372
Qy 481 CCCATTTAGCTCAAGTGTCTAAAGTTAATTAATTAATTAATTAATTAATTAAT 540
Db 371 TAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
Qy 541 GCAG 544
Db 311 AAAA 308

RESULT 15
US-10-239-676-132/c
Sequence 132, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: Olex, Alexander
APPLICANT: PEPENBERG, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 132
LENGTH: 5127
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (1976, 1981..1982, 1984, 1989..1990, 1995..1996, 5032, 5059)
US-10-239-676-132

Query Match 6.8%; Score 46.4; DB 5; Length 5127;
Best Local Similarity 47.9%; Fred. No. 1.9;
Matches 163; Conservative 0; Mismatches 176; Indels 1; Gaps 1;
Qy 184 AAAAAATTTCTCAAGAGCAATGGCCCAACAAAGCTCAAAAAATTAATTAATTAAT 243

Db	1173	AACATTTCTAAAAAATTTCTGGAAAAAATTAATCTCAAAAAAATTAATATTTT	1114
QY	244	AGAAAAAATCTGAATATATCAATCGTTAGAAVGGTGAATGAACTTACTGCGT	303
Db	1113	AAAAAACCAATCTCAAAAAAATTTTATTAATTTACGAACGAATATTTCAAAA	1054
QY	304	AATGGTATCTCAAGCAGAACCCCACTCTTTCTAAAACTTTGGCAATCAAGAA	363
Db	1053	AAAAAATTTTACCGAAAAAATTTTACCGAAAAAATTTCTAAAAAAT	994
QY	364	GAGTTGAATGCAAAACCTGGAAATTCATATACATTAACAATTAATGACTCAAAAAG-A	422
Db	993	TCCAAAAATATCTAATCTTAACAAAAAATTTCTAAAAAATTAATCAAAAAACGA	934
QY	423	ACCACTTATGCTAAAGTTAATAAAAAGAAACGAGGAGCAGAGCTAGCTTGAAGAAC	482
Db	933	AACCGAATAAAAAATTTCAAAAAAATTTCTAAATTAACAACCTTTAAAAAAA	874
QY	483	CATTAGCGCTCAAGTGTCAAAAAAGTAAATGCAAAATTT	522
Db	873	CATTTAAATTAATTAACATCATCTAAAAAATTAATAAACATTT	834

Search completed: March 6, 2006, 22:25:26
Job time : 786 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:38:13 ; Search time 183 Seconds
(without alignments)
6653.714 Million cell updates/sec

Title: US-09-360-685C-26

Perfect score: 685
Sequence: 1 aaaaatgcgaataaga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
 - 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
 - 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
 - 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
 - 5: /cgn2_6/ptodata/1/ina/6H COMB.seq:*
 - 6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq:*
 - 7: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
 - 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
 - 9: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	685	100.0	5925	3	US-08-470-260-4 Sequence 4, Appl1
2	685	100.0	5925	3	US-08-471-491-4 Sequence 4, Appl1
3	685	100.0	5925	3	US-08-466-662-4 Sequence 4, Appl1
4	685	100.0	10299	2	US-08-477-451-1 Sequence 1, Appl1
5	685	100.0	10299	2	US-08-477-451-1 Sequence 5, Appl1
6	685	100.0	19932	2	US-08-477-451-25 Sequence 25, Appl1
7	566.6	82.7	815	3	US-08-993-010-3 Sequence 3, Appl1
8	485.2	70.8	4821	2	US-08-053-614-3 Sequence 3, Appl1
9	485.2	70.8	4821	2	US-08-316-397B-3 Sequence 3, Appl1
10	485.2	70.8	4821	2	US-09-034-306-3 Sequence 3, Appl1
11	485.2	70.8	4821	2	US-09-259-437-3 Sequence 3, Appl1
12	485.2	70.8	4821	6	PCT-US93-09782-3 Sequence 3, Appl1
13	277.2	40.5	3648	2	US-08-053-614-1 Sequence 1, Appl1
14	277.2	40.5	3648	2	US-08-316-397B-1 Sequence 1, Appl1
15	277.2	40.5	3648	2	US-09-034-306-1 Sequence 1, Appl1
16	277.2	40.5	3648	2	US-09-259-437-1 Sequence 1, Appl1
17	277.2	40.5	3648	6	PCT-US93-09782-1 Sequence 1, Appl1
18	56	8.2	7218	2	US-08-232-463-14 Sequence 14, Appl1
19	45.6	6.7	849	3	US-09-248-796A-2496 Sequence 7183, Ap
20	44.8	6.5	1128	3	US-09-569-348-5 Sequence 5, Appl1
21	44.8	6.5	1131	9	US-08-795-475-5 Patent No. 5180810
22	44.8	6.5	1332	3	US-08-325-278B-5 Sequence 5, Appl1
23	44.8	6.5	1784	9	US-08-470-260-4 Patent No. 5180810
24	44.8	6.5	1784	9	US-08-470-260-4 Patent No. 5180810

25	44.4	6.5	612	3	US-09-902-540-1357 Sequence 1357, Ap
26	44	6.4	750	3	US-09-248-796A-499 Sequence 499, App
27	43.8	6.4	642	3	US-09-248-796A-4755 Sequence 4755, Ap
28	43.8	6.4	1039	3	US-09-902-540-1280 Sequence 1280, Ap
29	43.2	6.3	765	3	US-08-914-479A-3 Sequence 3, Appl1
30	43.2	6.3	918	3	US-08-937-271-9 Sequence 9, Appl1
31	42	6.1	606	3	US-09-248-796A-2496 Sequence 2496, Ap
32	41.6	6.1	855	3	US-08-914-479A-5 Sequence 5, Appl1
33	41.6	6.1	205163	3	US-09-949-016-17009 Sequence 17009, A
34	41.4	6.0	152070	3	US-09-949-016-155879 Sequence 155879, A
35	41.2	6.0	601	3	US-09-949-016-135996 Sequence 135996, A
36	41.2	6.0	601	3	US-09-949-016-136113 Sequence 136113, A
37	41.2	6.0	601	3	US-09-949-016-15553 Sequence 15553, A
38	41.2	6.0	193555	3	US-09-949-016-15554 Sequence 15554, A
39	41.2	6.0	193555	3	US-09-949-016-15554 Sequence 15555, A
40	41.2	5.9	6744	2	US-08-119-125A-2 Sequence 2, Appl1
41	40.6	5.9	30549	3	US-09-134-001C-322 Sequence 322, App
42	40.4	5.9	212	3	US-09-270-767-30271 Sequence 30271, A
43	40.2	5.9	1055	3	US-09-806-708B-23 Sequence 23, Appl1
44	40.2	5.9	1967	3	US-09-270-767-14158 Sequence 14158, A
45	40.2	5.9	1967	3	US-09-270-767-14158 Sequence 14158, A

ALIGNMENTS

RESULT 1
US-08-470-260-4
Sequence 4, Application US/08470260
Patent No. 6077706
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
TITLE OF INVENTION: for Vaccines and Diagnostics
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-470-260-4

Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2,4e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
DB AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
QY 2782 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
DB AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
QY 61 TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGTTAAAGTTGATTCATCA 120
DB TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGTTAAAGTTGATTCATCA 120
QY 121 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 180
DB GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 180
QY 2902 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 2961
DB GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 2961
QY 181 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 240
DB CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 240
QY 2962 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 3021
DB CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 300
DB GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 300
QY 3022 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 3081
DB GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 3081
QY 301 GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 360
DB GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 360
QY 3082 GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
DB GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 420
DB AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 420
QY 3142 AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3201
DB AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3201
QY 421 GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 480
DB GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 480
QY 3202 GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3261
DB GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3261
QY 481 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 540
DB CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 540
QY 3262 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 3321
DB CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 3321
QY 541 GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 600
DB GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 600
QY 3322 GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 3381
DB GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 3381
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 660
DB GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 660
QY 3382 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 3441
DB GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 3441
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
DB AATCTCAATCAAGCGGTATCAAGAG 685
QY 3442 AATCTCAATCAAGCGGTATCAAGAG 3466
DB AATCTCAATCAAGCGGTATCAAGAG 3466
```

RESULT 2
US-08-471-491-4
Sequence 4, Application US/08471491B
Patent No. 6090611

GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0044
CURRENT APPLICATION NUMBER: US/08/471,491B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 5925

TYPE: DNA
ORGANISM: Helicobacter pylori
US-08-471-491-4
Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2,4e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
DB AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 60
QY 2782 AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
DB AAAATGGCAAAATTAAGATTTCAGCAAGGTACGCAAGCAAAAGGACCTTGAAT 2841
QY 61 TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGTTAAAGTTGATTCATCA 120
DB TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGTTAAAGTTGATTCATCA 120
QY 121 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 180
DB GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 180
QY 2902 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 2961
DB GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAAAGGTAAAGGTTAAAG 2961
QY 181 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 240
DB CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 240
QY 2962 CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 3021
DB CTCAAAAATTTCTCAAGAGGCAATGGCCCAACAGCTCAAAAAATGAAGTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 300
DB GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 300
QY 3022 GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 3081
DB GCTAGAAAAAATCTGAATATATCAATCCGTTAAAGATGTGATGAATGAACTTCAAT 3081
QY 301 GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 360
DB GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 360
QY 3082 GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
DB GGTATGGGTTATCTCAAGAGGCAAGGCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
QY 361 AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 420
DB AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 420
QY 3142 AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3201
DB AAAGAGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3201
QY 421 GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 480
DB GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 480
QY 3202 GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3261
DB GAAAGGTTGAATGCAAACTTGGAAATTTGATTAACAATTAACAATTAATGAACTCAAAAAC 3261
QY 481 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 540
DB CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 540
QY 3262 CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 3321
DB CCCATTACGCTCAAGTTGCTTAAAGGTAAATGCAAAATTTGACGACTCATCAATA 3321
QY 541 GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 600
DB GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 600
QY 3322 GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 3381
DB GCAAGTGGTTGGGTTGTTGAGGCAAGCGGGCTTCCCTTTGAAAAGGATGATTA 3381
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 660
DB GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 660
QY 3382 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 3441
DB GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGGCTCAAGAAATTTGAC 3441
QY 661 AATCTCAATCAAGCGGTATCAAGAG 685
DB AATCTCAATCAAGCGGTATCAAGAG 685
QY 3442 AATCTCAATCAAGCGGTATCAAGAG 3466
DB AATCTCAATCAAGCGGTATCAAGAG 3466
```

RESULT 3
US-08-466-662-4
Sequence 4, Application US/0846662B
Patent No. 6130059

GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0057
CURRENT APPLICATION NUMBER: US/08/466,662B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 5925
TYPE: DNA
ORGANISM: Helicobacter pylori
US-08-466-662-4

Query Match 100.0%; Score 685; DB 3; Length 5925;
Best Local Similarity 100.0%; Pred. No. 2.4e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAAATTAAGATTTCACCAAGTAACCAAGCAAAAAGCCACTTGAAT 60
DB 2782 AAAAATGGCAAAAATTAAGATTTCACCAAGTAACCAAGCAAAAAGCCACTTGAAT 2841
QY 61 TCCGTTAAGATGATCATCATCAAAAAGTAACGATTAAGTATATCTCAATCA 120
DB 2842 TCCGTTAAGATGATCATCATCAAAAAGTAACGATTAAGTATATCTCAATCA 2901
QY 121 GCGGTATCAGTGGCTAAGCAAGGGTATTCAGTAGGGTGAAGCAAGCTTAAGCCGAT 180
DB 2902 GCGGTATCAGTGGCTAAGCAAGGGTATTCAGTAGGGTGAAGCAAGCTTAAGCCGAT 2961
QY 181 CTCAAAAATTTCTCAAAAGGCAATTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 240
DB 2962 CTCAAAAATTTCTCAAAAGGCAATTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 3021
QY 241 GCTAGAAAAAATCTGAATATATTCATCCGTTAAGATGTTGTAATGAAACCTTAGTC 300
DB 3022 GCTAGAAAAAATCTGAATATATTCATCCGTTAAGATGTTGTAATGAAACCTTAGTC 3081
QY 301 GGTAAATGGTTATCTCAAGCAAGGCAACTCTTTCTAAAACCTTTGCGACATCAAG 360
DB 3082 GGTAAATGGTTATCTCAAGCAAGGCAACTCTTTCTAAAACCTTTGCGACATCAAG 3141
QY 361 AAAGATTTGAATGCAAAAATCTGAAATTTCAATTAACAATTAATGACTCAAAAAC 420
DB 3142 AAAGATTTGAATGCAAAAATCTGAAATTTCAATTAACAATTAATGACTCAAAAAC 3201
QY 421 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGA 480
DB 3202 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGA 3261
QY 481 CCCATTTAGCTCAAGTGTCTAAAAAGTAAATGCAAAAATTTGACCGACTCAATCAATA 540
DB 3262 CCCATTTAGCTCAAGTGTCTAAAAAGTAAATGCAAAAATTTGACCGACTCAATCAATA 3321
QY 541 GCAAGTGTGTTGGTGTGTGTAAGGCGCAAGCGGGCTTCCCTTTGAAAAGGCAATGAA 600
DB 3322 GCAAGTGTGTTGGTGTGTGTAAGGCGCAAGCGGGCTTCCCTTTGAAAAGGCAATGAA 3381
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAAATGAC 660
DB 3382 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAAATGAC 3441
QY 661 AATTCATCAAGCCGATCAAG 685
DB 3442 AATTCATCAAGCCGATCAAG 3466
```

RESULT 4
US-08-477-451-1

Sequence 1, Application US/08477451
Patent No. 592865
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSES: Chilton Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-477-451-1

Query Match 100.0%; Score 685; DB 2; Length 10299;
Best Local Similarity 100.0%; Pred. No. 2.8e-158;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAATGGCAAAAATTAAGATTTCAGCAAGTAACCAAGCAAAAAGCCACTTGAAT 60
DB 7156 AAAAATGGCAAAAATTAAGATTTCAGCAAGTAACCAAGCAAAAAGCCACTTGAAT 7215
QY 61 TCCGTTAAGATGATCATCATCAAAAAGTAACGATTAAGTATATCTCAATCA 120
DB 7216 TCCGTTAAGATGATCATCATCAAAAAGTAACGATTAAGTATATCTCAATCA 7275
QY 121 GCGGTATCAGTGGCTAAGCAAGGGTATTCAGTAGGGTGAAGCAAGCTTAAGCCGAT 180
DB 7276 GCGGTATCAGTGGCTAAGCAAGGGTATTCAGTAGGGTGAAGCAAGCTTAAGCCGAT 7335
QY 181 CTCAAAAATTTCTCAAAAGGCAATTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 240
DB 7336 CTCAAAAATTTCTCAAAAGGCAATTGGCCCAACAAGCTCAAAAAATGAAGTCTCAAT 7395
QY 241 GCTGAAAAAATCTGAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGT 300
DB 7396 GCTGAAAAAATCTGAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGT 7455
QY 301 GGTAAATGGTTATCTCAAGCAAGGCAACTCTTTCTAAAACCTTTGCGACATCAAG 360
DB 7456 GGTAAATGGTTATCTCAAGCAAGGCAACTCTTTCTAAAACCTTTGCGACATCAAG 7515
QY 421 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGAA 480
DB 7516 GAACCATTTTATGCTAAATTAATTAAGAAAGCAAGGCGCAAGCTAGCCTTGAAGAA 7575
QY 481 CCCATTTAGCTCAAGTGTCTAAAAAGTAAATGCAAAAATTTGACCGACTCAATCAATA 540
DB 7576 CCCATTTAGCTCAAGTGTCTAAAAAGTAAATGCAAAAATTTGACCGACTCAATCAATA 7635
QY 541 GCAAGTGTGTTGGTGTGTGTAAGGCGCAAGCGGGCTTCCCTTTGAAAAGGCAATGAA 600
DB 7636 GCAAGTGTGTTGGTGTGTGTAAGGCGCAAGCGGGCTTCCCTTTGAAAAGGCAATGAA 7695
QY 601 GTTATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAAATGAC 660
```

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rtf

Page 4

Dd	7756	GTTCATGATCTCAGTAAAGTAGGGCTTTCAAGAAATCAAGAAATGGCTCGAATAATTAC	7815
Qy	661	AATCTCAATCAAGCGGTATCAGAG	685
Dd	7816	AATCTCAATCAAGCGGTATCAGAG	7840

RESULT 5
US-08-47

Sequence 5, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: C...

/ TITLE OF INVENTION: Helicobacter Pylori CagI Region
 / NUMBER OF SEQUENCES: 46
 / CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA
ZIP: 94608-29
COMPUTER READABLE

MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS 3.3
SOFTWARE: DataStar 1.0

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,451
 FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 3311

REFERENCE NUMBER: 33,113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708

TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10299 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
S-08-477-451-5
Query Match 100.0%

```

Best Local Similarity    100.0%; Score 685; DB 2; Length 10299;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

1 AAAAAATGCGAAAAATAGGATTTCAGCAGAGTAAACGACGAAAAACGACCTTGAAT 60
3144 AAAAAATGCGAAAAATAGGATTTCAGCAGAGTAAACGACGAAAAACGACCTTGAAT 208

61 TCCGTTAAGATGTGATCATCAATCAAAGGTAAACGATAAGTGTATCTCAATCAA 120
|||||
|||
3084 TCCGTTAAGATGTGATCATCAATCAAAGGTAAACGATAAGTGTATCTCAATCAA 120
|||||

121 GCGGTATCATGTGCTAAAGCAACGGGTGATTTCACTAGGGTAGACCAAGCTTAGCCGAT 180
|||||
3024 GCGGTATCATGTGCTAAAGCAACGGGTGATTTCACTAGGGTAGACCAAGCTTAGCCGAT 180
|||||

181 CTCAAAATTTCTCAAGAGCAATTGGCCACAGCTCAAAAAATGAAAGTCTCAAT 240
2064

241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGCTGTAATGCAACCTAGTTC 300
2504 CTCAAAAATTTCTCAAGAGACCAATTGGCCCMCAAGCTCAAAAAAATGAAAGTCTCAAT 2905

2904 GCTAGAAAAAATCTGAATAATATCATTCGTTAAGAATGTGTGAATGGAAACCTACTC 2845

301 GGTAATGGATTATCTCAAGCAGAACCAACAATTCTTCTAAAATCGTTTTGGGATCCGAT 2846

360 CAGCATCAG

Db 2844 GGTAATGGGTATCTCGACAGCAACCAACCTTTCTTAAAACTTTGGACATCAAG 2785
QY 361 AAAGAGTTAATGCAAACTTGAAATTTTCAATTAACAATACATTAATGACTGCAAAAC 420
Db 2784 AAAGAGTTAATGCAAACTTGAAATTTTCAATTAACAATACATTAATGACTGCAAAAC 420
QY 421 GAACCCATTTATGCTTAAGTTATATTAAGAAGAGAGGACAGACAGCTACCTTGAGAA 2725
Db 2724 GAACCCATTTATGCTTAAGTTATATTAAGAAGAGAGGACAGACAGCTACCTTGAGAA 2665
QY 481 CCCATTTACGCTCAAGTCTCTAAAAGATTAATGCAAAAATTGACCCAGCTCAATCAATA 540
Db 2664 CCCATTTACGCTCAAGTCTCTAAAAGATTAATGCAAAAATTGACCCAGCTCAATCAATA 2805
QY 541 GCAAAGTGTGTGGTGTGTATGAGGCAAGCAGGGGCTCTCCCTTGAAAAGAGCATTA 600
Db 2604 GCAAAGTGTGTGGTGTGTATGAGGCAAGCAGGGGCTCTCCCTTGAAAAGAGCATTA 2545
QY 601 GTTATTCATCTCAGTAAGGTAGGGCTTTCAAGCAATCAAGAAATGCTCAGAAAATTGAC 660
Db 2544 GTTATTCATCTCAGTAAGGTAGGGCTTTCAAGCAATCAAGAAATGCTCAGAAAATTGAC 2485
QY 661 AATCTCAATCAAGCCGATCAAGAG 685
Db 2484 AATCTCAATCAAGCCGATCAAGAG 2460

RESULT 6
US-08-477-451-25

sequence 25, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:
APPLICANT: Covaccl Antenn-11

TITLE OF INVENTION: Helicobacter Pylori CagI Region
 NUMBER OF SEQUENCES: 46
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Datatrak Database "v. 1.0"

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 32113

REFERENCE/DOCKET NUMBER: 0335.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELETYPE: 510-601-2708

TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 19932 base pairs

```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

MOLECULE TYPE: DNA (genomic)
S-08-477-451-25
Query Match 100.0%

Best Local Similarity	100.0%;	Score 685;	DB 2;	Length 19932;
Matches 685;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Pred. No.	3.3e-158;			

1 AAAAATGGCAAAATAAGCATTTACGAAAGTAAACGCAAAAGCGACCTTGAAAT 60

Db 16789 AAAAATGGCAAAATTAAGATTTCAGACAGGTAAAGCAAGCAAAAGCACTTGAAT 16848
 Qy 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGGTAAAGTGAATATCTCAATCAA 120
 Db 16849 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGGTAAAGTGAATATCTCAATCAA 16908
 Qy 121 GCGGTATCACTGCTAAAGCAAGCGGTGATTTAGTAGGTAGAGCAAGCGGTAGCCGAT 180
 Db 16909 GCGGTATCACTGCTAAAGCAAGCGGTGATTTAGTAGGTAGAGCAAGCGGTAGCCGAT 16968
 Qy 181 CTCAAAAATTTCTCAAAAGGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
 Db 16969 CTCAAAAATTTCTCAAAAGGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCAAT 17028
 Qy 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAAGATGTGTGAATGTGAACCTTAGTC 300
 Db 17029 GCTAGAAAAAATCTGAATATATCAATCGTTAAAGATGTGTGAATGTGAACCTTAGTC 17088
 Qy 301 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTTCTAAAAAATTTTGGACATCAAG 360
 Db 17089 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTTCTAAAAAATTTTGGACATCAAG 17148
 Qy 361 AAAGATGTGAATGCAAAACTTGGAATTTCAATPAACAATAATGACTCAAAAC 420
 Db 17149 AAAGATGTGAATGCAAAACTTGGAATTTCAATPAACAATAATGACTCAAAAC 17208
 Qy 421 GAACCAATTAATGCTAAAGTTATTAAGAAAGAGCGGCAACAGCTAGCTTGAAGAA 480
 Db 17209 GAACCAATTAATGCTAAAGTTATTAAGAAAGAGCGGCAACAGCTAGCTTGAAGAA 17268
 Qy 481 CCCATTTAAGCTCAAGCTTGAAGTTATTAAGAAAGAGCGGCAACAGCTAGCTTGAAGAA 540
 Db 17269 CCCATTTAAGCTCAAGCTTGAAGTTATTAAGAAAGAGCGGCAACAGCTAGCTTGAAGAA 17328
 Qy 541 GCAAGCTGTTGGGTGTGTAAGGCAAGCGGCTTCCCTTGAAGAAAGCAATGATAA 600
 Db 17329 GCAAGCTGTTGGGTGTGTAAGGCAAGCGGCTTCCCTTGAAGAAAGCAATGATAA 17388
 Qy 601 GTTATGATCTCAGTAAGGTAAGGCTTTCAAGAAATCAAGAAATTTGCTCAAGAAATTTGAC 660
 Db 17389 GTTATGATCTCAGTAAGGTAAGGCTTTCAAGAAATCAAGAAATTTGCTCAAGAAATTTGAC 17448
 Qy 661 AATCTCAATCAAGCGGTATCAAG 685
 Db 17449 AATCTCAATCAAGCGGTATCAAG 17473

RESULT 7
 US-08-993-010-3
 / Sequence 3, Application US/08993010
 / Patent No. 6902903
 / GENERAL INFORMATION:
 / APPLICANT: Quan, Stella
 / APPLICANT: Valenzuela, Pablo
 / APPLICANT: Polico, Alan
 / TITLE OF INVENTION: HELICOBACTER PYLORI DIAGNOSTICS
 / FILE REFERENCE: 1360 002
 / CURRENT APPLICATION NUMBER: US/08/993,010
 / EARLIER FILING DATE: 1997-12-18
 / EARLIER APPLICATION NUMBER: 60/033,707
 / NUMBER OF SEQ ID NOS: 4
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 3
 / LENGTH: 815
 / TYPE: DNA
 / ORGANISM: Helicobacter pylori
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (1)..(804)
 US-08-993-010-3

Query Match 82.7%; Score 566.6; DB 3; Length 815;

Best Local Similarity 94.2%; Pred. No. 1.5e-129;
 Matches 613; Conservative 0; Mismatches 29; Indels 9; Gaps 2;
 Qy 1 AAAAATGGCAAAATTAAGATTTCAGACAGGTAAAGCAAGCAAAAGCACTTGAAT 60
 Db 10 AAAAATGGCAAAATTAAGATTTCAGACAGGTAAAGCAAGCAAAAGCACTTGAAT 69
 Qy 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGGTAAAGTGAATATCTCAATCAA 120
 Db 70 TCCGTTAAAGATGATCATCAATCAAAAGGTAAAGGTAAAGTGAATATCTCAATCAA 129
 Qy 121 GCGGTATCACTGCTAAAGCAAGCGGTGATTTAGTAGGTAGAGCAAGCGGTAGCCGAT 180
 Db 130 GCGGTATCACTGCTAAAGCAAGCGGTGATTTAGTAGGTAGAGCAAGCGGTAGCCGAT 189
 Qy 181 CTCAAAAATTTCTCAAAAGGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCAAT 240
 Db 190 CTCAAAAATTTCTCAAAAGGCAATTTGCCCAACAGCTCAAAAAATGAAGTCTCAAT 249
 Qy 241 GCTAGAAAAAATCTGAATATATCAATCGTTAAAGATGTGTGAATGTGAACCTTAGTC 300
 Db 250 GCTAGAAAAAATCTGAATATATCAATCGTTAAAGATGTGTGAATGTGAACCTTAGTC 309
 Qy 301 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTTCTAAAAAATTTTGGACATCAAG 360
 Db 310 GGTAAATGGTTATCTCAAGCAAGCAACCTCTTTCTAAAAAATTTTGGACATCAAG 369
 Qy 361 AAAGATGTGAATGCAAAACTTGGAATTTCAATPAACAATAATGACTCAAAAC 417
 Db 370 AAAGATGTGAATGCAAAACTTTTGAATTTCAATPAACAATAATGACTCAAAAC 429
 Qy 418 AAC-----GAACCAATTAATGCTAAAGTTATTAAGAAAGAGCGGCAACAGCTAGC 471
 Db 430 AACAGCAAGAACCCATTTATCTAAAGTTATTAAGAAAGAGCGGCAACAGCTAGC 489
 Qy 472 CTTGAAGAAACCATTTAGCTCAAGTTGCTAAAGAAAGGTAAAGTGAACGACTC 531
 Db 490 CTTGAAGAAACCATTTATATCTAAAGTTGCTAAAGAAAGGTAAAGTGAACGACTC 549
 Qy 532 AATCAATTAAGCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTTCCCTTGAAGAAAG 591
 Db 550 AATCAATTAAGCAAGTGTGTTGGGTGTGTAAGGCAAGCGGCTTCCCTTGAAGAAAG 609
 Qy 592 CATGATTAAGTTGATCTCAGTAAGGTAAGGCTTTCAAGAAATCAAGAA 642
 Db 610 CATGATTAAGTTGATCTCAGTAAGGTAAGGCGCATGATGACCTGAA 660

RESULT 8
 US-08-053-614-3
 / Sequence 3, Application US/08053614
 / Patent No. 5403924
 / GENERAL INFORMATION:
 / APPLICANT: COVER, TIMOTHY L.
 / APPLICANT: BLASER, MARTIN J.
 / APPLICANT: TUMMURU, MURALI K.R.
 / TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
 / NUMBER OF SEQUENCES: 4
 / CORRESPONDENCE ADDRESS:
 / ADDRESS: NEEDLE & ROSENBERG, P.C.
 / STREET: 133 Carnegie Way, Suite 400
 / CITY: Atlanta
 / STATE: Georgia
 / COUNTRY: USA
 / ZIP: 30303
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: PatentIn Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/053,614

Tue Mar 7 12:58:26 2006

us-09-360-685c-26.rn1

Page 8

```

/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/316,397
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SPRATT, GWENDOLYN D
/ REGISTRATION NUMBER: 36,016
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 404/688-0770
/ TELEFAX: 404/688-9880
/ INFORMATION FOR SEQ ID NO:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4821 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1072..4614
/
/ US-09-259-437-3

```

Query Match
Best Local Similarity 70.8%; Score 485.2; DB 3; Length 4821;
Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

```

/ 1 AAAAAAGCAAAATTAAGATTTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 60
/ 3319 AAAAAAGCAAAATTAAGATTTTTCAGCAAGTAAAGCAAGCAAAAGGACCAAGAAAT 3378
/ 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCA 120
/ 3379 TCATTAAGATGTATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCA 3438
/ 121 GCGGTATCAGTGTCTAAAGCAAGGATTTTCAGTAAAGGATTAAGTTGATTAATCTCA 180
/ 3439 GCGGTATCAGTGTCTAAAGCAAGGATTTTCAGTAAAGGATTAAGTTGATTAATCTCA 3498
/ 181 CTCAAAAATTTCTCAAGAGCAAGTGGCCCAAGCTCAAAAGTAAAGTTGATTAATCTCA 240
/ 3499 CTCAAAAATTTCTCAAGAGCAAGTGGCCCAAGCTCAAAAGTAAAGTTGATTAATCTCA 3558
/ 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATGTGAATGGAACCTTAGTC 300
/ 3559 GTTGG---AAATCTGAATATATCAATCCGTTAAGATGTGAATGGAACCTTAGTC 3615
/ 301 GGTATGGTTATCTCAAGCAAGCAAGCAAGCTTTCTAAAGCTTTTGGACATCAAG 360
/ 3616 GGTATGGTTATCTCAAGCAAGCAAGCAAGCTTTCTAAAGCTTTTGGACATCAAG 3675
/ 361 AAAGATTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGACTCAAAAC 420
/ 3676 AAAGATTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGACTCAAAAC 474
/ 421 -----GAACCAATTAATGCTTAAGTTAAATTAAGAAAGAGGCAAGCTTAGCTT 3735
/ 3736 GCGGAGAGCAACCAATTAATGCTTAAGTTAAATTAAGAAAGAGGCAAGCTTAGCTT 3795
/ 475 GAAGACCAATTAAGCTTAAGTTGCTTAAGAAAGGATTAATGCAAAATTTGACCTCAAT 534
/ 3796 GAAGACCAATTAAGCTTAAGTTGCTTAAGAAAGGATTAATGCAAAATTTGACCTCAAT 3855
/ 535 CAATATGCAAGTGGTTGGGCTTTGAGGCAAGCAAGGAGGCTTCCCTTGAAGAGGCAAT 594
/ 3856 CAAGCAAGCAAGTGGTTGGGCTTTGAGGCAAGCAAGGAGGCTTCCCTTGAAGAGGCAAT 3915
/ 595 GATTAAGTTGATGATCTCAAGCAAGCAAGCAAGCTTTCTAAAGATCAAGAA 642
/ 3916 GATTAAGTTGATGATCTCAAGCAAGCAAGCAAGCTTTCTAAAGATCAAGAA 3963

```

RESULT 12

```

/ PCT-US93-09782-3
/ Sequence 3, Application PC/TUS9309782
/ GENERAL INFORMATION:
/ APPLICANT: COVER, TIMOTHY L.
/ APPLICANT: BLASER, MARTIN J.
/ APPLICANT: TUMMURU, MURALI K. R.
/ TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: NEEDLE & ROSENBERG, P.C.
/ STREET: 133 Carnegie Way, Suite 400
/ CITY: Atlanta
/ STATE: Georgia
/ COUNTRY: USA
/ ZIP: 30303
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/09782
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SPRATT, GWENDOLYN D
/ REGISTRATION NUMBER: 36,016
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 404/688-0770
/ TELEFAX: 404/688-9880
/ INFORMATION FOR SEQ ID NO:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4821 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1072..4614
/
/ PCT-US93-09782-3

```

Query Match
Best Local Similarity 70.8%; Score 485.2; DB 6; Length 4821;
Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

```

/ 1 AAAAAAGCAAAATTAAGATTTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 60
/ 3319 AAAAAAGCAAAATTAAGATTTTTCAGCAAGTAAAGCAAGCAAAAGGACCAAGAAAT 3378
/ 61 TCCGTTAAAGATGTATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCA 120
/ 3379 TCATTAAGATGTATCATCAATCAAAAGTAAAGGATTAAGTTGATTAATCTCA 3438
/ 121 GCGGTATCAGTGTCTAAAGCAAGGATTTTCAGTAAAGGATTAAGTTGATTAATCTCA 180
/ 3439 GCGGTATCAGTGTCTAAAGCAAGGATTTTCAGTAAAGGATTAAGTTGATTAATCTCA 3498
/ 181 CTCAAAAATTTCTCAAGAGCAAGTGGCCCAAGCTCAAAAGTAAAGTTGATTAATCTCA 240
/ 3499 CTCAAAAATTTCTCAAGAGCAAGTGGCCCAAGCTCAAAAGTAAAGTTGATTAATCTCA 3558
/ 241 GCTAGAAAAAATCTGAATATATCATCCGTTAAGATGTGAATGGAACCTTAGTC 300
/ 3559 GTTGG---AAATCTGAATATATCAATCCGTTAAGATGTGAATGGAACCTTAGTC 3615
/ 301 GGTATGGTTATCTCAAGCAAGCAAGCAAGCTTTCTAAAGCTTTTGGACATCAAG 360
/ 3616 GGTATGGTTATCTCAAGCAAGCAAGCAAGCTTTCTAAAGCTTTTGGACATCAAG 3675
/ 361 AAAGATTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGACTCAAAAC 420

```

Db 3676 AAGAAATGATGAGAAATTAATAATTCAATACATACATATATGCTCAAAAAC 3735
Qy 421 -----GACCCTTTATGCTAAAGTAAATAAAAAGACAGGCGACAGCTACCTT 474
Db 3736 GCGCGAGAACCCCTTTATGCTCAAGTTAAATAAAAAGACAGACAGTAGCTACCCCT 3795
Qy 475 GAAGAACCATTTACCTCAAGTTGCTAAATAGTAAATGACCAAGCTCAAT 534
Db 3796 GAAGAACCATTTATGCTCAAGTTGCTAAATAGTAAATGACCAAGCTCAAT 3855
Qy 535 CAATAGCAAGTGGTTGGTGTGTGAAGGCAAGCGGCTTCCCTTGAAGAGCAT 594
Db 3856 CAAGCAGCAGCAAGTGGTTGGTGTGTGAAGGCAAGCGGCTTCCCTTGAAGAGCAT 3915
Qy 595 GATTAAGTATGATCTCAGTAAGGTAAGGCTTTCAGAGATCAAGA 642
Db 3916 GATTAAGTATGATCTCAGTAAGGTAAGGCTTTCAGAGATCAAGA 3963

RESULT 13

US-08-053-614-1
Sequence 1, Application US/08053614
Patent No. 5403924
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TOMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3648 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..3648
US-08-053-614-1

Query Match 40.5%; Score 277.2; DB 2; Length 3648;
Best Local Similarity 92.1%; Pred. No. 2e-58;
Matches 304; Conservative 0; Mismatches 23; Indels 3; Gaps 1;
Qy 1 AAAAAATGCGAAAAATATGAGATTTCAGCAAGTACGCAAGCAAAAAAGCAAGCTTGAAT 60

Db 3319 AAAAAATGCGAAAAATATGAGATTTCAGCAAGTACGCAAGCAAAAAAGCAAGCAAGAAAT 3378
Qy 61 TCCGTTAAAGATGTGATCATCATCAAAAGGTAAAGGTATATCTCAATCA 120
Db 3379 TCCATTAAGATGTGATCATCATCAAAAGGTAAAGGTATATCTCAATCA 3438
Qy 121 GGGGTATCAGTGGCTTAAGCAAGGGTATTTCAAGGGTATGAGCAAGCTTACCGAT 180
Db 3439 GGGGTATCAGTGGCTTAAGTATACCGTGCATTTCAAGGGTATGAGCAAGCTTACCGAT 3498
Qy 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCAAT 240
Db 3499 CTCAAAAATTTCTCAAGAGCAATTTGCTCAACAAAGCTCAAAAAATGAAGTTCAT 3558
Qy 241 GCTAGAAAAAAATCTGAATATATCATCCGTTAAGAAATGTGTGAATGAACCTAGTC 300
Db 3559 GTTGG--AAATCTGAATATATACCAATCCGTTAAGAAATGTGTGAACCGAACCTAGTC 3615
Qy 301 GGTAAATGGTTATCTCAAGCAGACGCCACA 330
Db 3616 GGTAAATGGTTATCTCAAGATAGGCCACA 3645

RESULT 14

US-08-316-397B-1
Sequence 1, Application US/08316397B
Patent No. 5733740
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TOMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3648 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: 1072..3648
US-08-316-397B-1

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 6, 2006, 21:25:19 ; Search time 555 Seconds

(without alignments)
8225.792 Million cell updates/sec

Title: US-09-360-685C-26

Sequence: 1 aaatagcgaataagga.....caatcaagcgatcagaag 685

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_21.*
2: geneseqn1908.*
3: geneseqn2000.*
4: geneseqn2001.*
5: geneseqn2002.*
6: geneseqn2003.*
7: geneseqn2004.*
8: geneseqn2005.*
9: geneseqn2006.*
10: geneseqn2007.*
11: geneseqn2008.*
12: geneseqn2009.*
13: geneseqn2010.*
14: geneseqn2011.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	100.0	4149	2	AA62461 Helicobac
2	685	100.0	5925	2	AA62461 Helicobac
3	685	100.0	19932	2	AA62461 Helicobac
4	531.6	77.6	3558	13	ADU05406 DNA encod
5	502.6	73.4	1024	6	ABX66359 Helicobac
6	497.6	72.6	1024	6	ABX66359 Helicobac
7	485.2	70.8	4821	2	AA62461 Helicobac
8	485.2	70.8	4821	2	AA62461 Helicobac
9	485.2	70.8	4821	2	AA62461 Helicobac
10	482	70.4	3546	3	AA62461 Helicobac
11	477.2	69.7	3546	3	AA62461 Helicobac
12	477.2	69.7	3546	3	AA62461 Helicobac
13	477.2	69.7	3546	3	AA62461 Helicobac
14	406.2	59.3	1024	6	ABX66359 Helicobac
15	379.4	55.4	854	6	ABX66359 Helicobac
16	379.4	55.4	1024	6	ABX66359 Helicobac
17	379.4	55.4	1024	6	ABX66359 Helicobac
18	277.2	40.5	3648	2	AA62461 Helicobac
19	277.2	40.5	3648	2	AA62461 Helicobac

20	271.2	39.6	762	6	ABX66359 Helicobac
21	268.4	39.2	1008	6	ABX66359 Helicobac
22	48.8	7.1	1024	6	ABX66359 Helicobac
23	48	7.0	12138	6	ABX66359 Helicobac
24	48	7.0	12138	6	ABX66359 Helicobac
25	47.6	6.9	863	4	AA195036 DNA trans
26	47	6.9	14006	6	ABX66359 Helicobac
27	46.8	6.8	5542	6	ABX66359 Helicobac
28	46.8	6.8	9964	6	ABX66359 Helicobac
29	46.8	6.8	10205	6	ABX66359 Helicobac
30	46.8	6.8	10205	6	ABX66359 Helicobac
31	46.4	6.8	5127	4	AA545425 Chemical
32	46	6.7	7969	6	ABX66359 Helicobac
33	46	6.7	7969	6	ABX66359 Helicobac
34	45.6	6.7	6182	6	ABX66359 Helicobac
35	45.4	6.6	14362	6	ABX66359 Helicobac
36	45.4	6.6	14708	4	AA545514 Chemical
37	45.4	6.6	14708	4	ABX66359 Helicobac
38	45.4	6.6	14708	6	ABX66359 Helicobac
39	45.4	6.6	14708	6	ABX66359 Helicobac
40	45.2	6.6	8222	8	ACF62794 Colon can
41	45.2	6.6	8222	8	ACF62794 Colon can
42	45.2	6.6	11222	10	ADB54190 Pretreat
43	45.2	6.6	11222	10	ADB54190 Pretreat
44	45	6.6	5814	6	ABX66359 Helicobac
45	45	6.6	9666	6	ABX66359 Helicobac

ALIGNMENTS

RESULT 1	AA62461	standard; DNA; 4149 BP.
ID	AA62461	standard; DNA; 4149 BP.
XX	AA62461	
AC	AA62461	
XX	AA62461	
DT	17-OCT-2003	(revised)
DT	15-FEB-1999	(first entry)
XX		
DE	Helicobacter pylori caga-Vibrio cholerae A2+B subunit gene fusion.	
XX		
KM	Vaccine; antigen; CagB; antigen; toxin; diagnosis; gastritis; ulcer;	
KM	stomach cancer; ss.	
XX		
OS	Helicobacter pylori.	
OS	Vibrio cholerae.	
OS	Chimeric.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..3444
FT		/*tag= a
FT		/product= "CagA"
FT		3444..4020
FT		/*tag= b
FT		/product= "A2+B subunits"
XX		
XX	W09844130-A1.	
XX		
PD	08-OCT-1998.	
XX		
PF	31-MAR-1998;	98WO-KR000073.
XX		
PR	31-MAR-1997;	97KR-00011950.
XX		
XX	31-MAR-1997;	97KR-00011951.
XX		
PA	(DAEW-) DAEWONG PHARM CO LTD.	
XX		
PI	Kim B, Shin S, Yu Y, Park M, Choi D, Jung H;	
XX		
DR	WPI, 1998-568279/48.	
XX	P-PsDB; AAW80600.	

PT New chimeric proteins for use against *Helicobacter pylori* - comprising an
 PT antigenic protein of *H. pylori* and A1 and B subunits of Vibrio cholerae
 XX toxin, preferably produced by recombinant techniques.
 PS Claim 4; Page 78-80; 102pp; English.

XX This is the nucleotide sequence of a fusion gene prepared by ligating the
 CC cagA gene of *Helicobacter pylori* and the A2 and B subunits genes of
 CC Vibrio cholerae toxin. It codes for a claimed 1338-amino acid fusion
 CC protein (see AWW86000). The invention relates to: chimeric proteins
 CC comprising antigenic proteins of *H. pylori* and A2 and B subunits of *V.*
 CC cholerae toxin; recombinant DNAs encoding such chimeric proteins;
 CC recombinant expression vectors; a process for preparing the chimeric
 CC proteins using transformed microbial host cells; and to preventative and
 CC therapeutic vaccines comprising the chimeric proteins for *H. pylori*-
 CC associated diseases such as gastritis, gastric ulcer, duodenal ulcer and
 CC gastric cancer (claimed). The chimeric proteins are designed to have
 CC excellent immunogenicity, to be stable in the stomach, to penetrate the
 CC mucous membrane of the intestine, and to stimulate production of sIgA.
 CC They can additionally be used as active ingredients in diagnostic kits
 CC for *H. pylori* infection, and for production of anti-*H. pylori* antibody.
 CC (Updated on 17-Oct-2003 to standardise OS field)

Sequence 4149 BP; 1544 A; 714 C; 819 G; 1072 T; 0 U; 0 Other;

Query Match Best Local Similarity 100.0%; Score 685; DB 2; Length 4149;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAGCAAAATTAAGATTTTCAGACAGTACGACGCAAAAGACCTTGAAT 60
 DB 2248 AAAAAAGCAAAATTAAGATTTTCAGACAGTACGACGCAAAAGACCTTGAAT 60
 QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAAGTAAAGTAAATCTCAATCA 120
 DB 2308 TCCGTTAAAGATGTGATCATCAATCAAAAGTAAAGTAAAGTAAATCTCAATCA 120
 QY 121 GCGGTATCAGTGGCTAAAGCAACGGGTATTCAGTGGGTAGACAGCGTTAGCCGAT 180
 DB 2368 GCGGTATCAGTGGCTAAAGCAACGGGTATTCAGTGGGTAGACAGCGTTAGCCGAT 180
 QY 181 CTCAAAAATTTCTCAAGAGCAATTCGCGCAACACTCAAAAAATGAAGTCTCAAT 240
 DB 2428 CTCAAAAATTTCTCAAGAGCAATTCGCGCAACACTCAAAAAATGAAGTCTCAAT 240
 QY 241 GCTAGAAAAAATTCGAATATATCAATCCGTTAAAGTGTGATGAGACCTTATGTC 300
 DB 2488 GCTAGAAAAAATTCGAATATATCAATCCGTTAAAGTGTGATGAGACCTTATGTC 300
 QY 301 GGTATGCGTATCTCAAGAGCAACCTTTCTTAAATCTTTGGACATCAAG 360
 DB 2548 GGTATGCGTATCTCAAGAGCAACCTTTCTTAAATCTTTGGACATCAAG 360
 QY 361 AAGAGTTGAATGCAAACTTGAATTTCAATCAATCAATTAATGAGACTCAAAAC 420
 DB 2608 AAGAGTTGAATGCAAACTTGAATTTCAATCAATCAATTAATGAGACTCAAAAC 420
 QY 421 GAACCATTTATGCTAAAGTAAATTAAGAAAGAGCGGCAAGAGCTTGAAGAA 480
 DB 2668 GAACCATTTATGCTAAAGTAAATTAAGAAAGAGCGGCAAGAGCTTGAAGAA 480
 QY 481 CCCATTACGCTCAAGTCTTAAAGGTAAATGCAAAATTTGACGACTCAATATA 540
 DB 2728 CCCATTACGCTCAAGTCTTAAAGGTAAATGCAAAATTTGACGACTCAATATA 540
 QY 541 GGAAGTGGTTGGGTGTGAGGCAAGACGGGCTTCCTTTGAAAAGGAGATATA 600
 DB 2788 GGAAGTGGTTGGGTGTGAGGCAAGACGGGCTTCCTTTGAAAAGGAGATATA 600
 QY 601 GTTATGATCTCAGTAAAGGTAGGCTTTCAAGGAATCAAGATTTGGCTAGAAAAATGAC 660
 DB 2848 GTTATGATCTCAGTAAAGGTAGGCTTTCAAGGAATCAAGATTTGGCTAGAAAAATGAC 660
 DB 2907 GTTATGATCTCAGTAAAGGTAGGCTTTCAAGGAATCAAGATTTGGCTAGAAAAATGAC 2907

QY 661 AATTCATCAAGCGGTATCAAG 695
 DB 2908 AATTCATCAAGCGGTATCAAG 2932

RESULT 2
 AA048733
 ID AA048733 standard; DNA; 5925 BP.
 AC AA048733;
 DN 25-MAR-2003 (revised)
 DT 17-MAR-1994 (first entry)
 XX CAI gene.
 XX Cytotoxin; CT; *H. pylori*; precursor; vacuolation; cell death; eukaryote;
 KW cytotoxin-associated immunodominant antigen; CAI heat shock protein;
 KW hsp60; type B; gastritis; peptic ulcer; gastric tumours; ss.
 XX *Helicobacter pylori*.
 OS
 XX
 XX Key Location/Qualifiers
 FH RBS 525..529
 FT /*cag= b
 FT CDS 535..3978
 FT /*cag= a
 FT /product= "CAI antigen"
 FT terminator 4010..4032
 FT /*cag= c
 FT /standard_name= "Rho_independent_terminator"
 XX
 XX MO9318150-A1.
 XX 16-SEP-1993.
 XX
 XX 02-MAR-1993; 93MO-BP000472.
 XX
 XX 02-MAR-1992; 92IT-PI000052.
 PR 25-JAN-1993; 93MO-BP000158.
 XX
 XX (ISTS) BIOICNE SCLAVO SPA.
 PA
 XX
 XX Covacci A, Bugnoli M, Telford J, Macchia G, Rappuoli R;
 DR WPI: 1993-303464/38.
 DR P-PSDB; AAR41199.
 XX
 PT Recombinant *Helicobacter pylori* protein and coresep. gene - is a
 PT cytotoxin, antigen or heat shock protein used for treating and preventing
 PT type B gastritis, gastric ulcers and gastric tumours.
 XX
 PS Claim 28; Fig 4; 83pp; English.
 XX
 CC This sequence encodes the cytotoxin-associated immunodominant antigen,
 CC CAI, of *H. pylori*. The protein is a hydrophilic, surface-exposed protein
 CC having a molecular weight of approximately 120-132 kD. Pref 128-150 kD,
 CC and an isoelectric point of 9.72. *H. pylori* isolates which do not
 CC produce the CAI antigen do not have the *cag* gene, and are also unable to
 CC produce an active cytotoxin. The association between the presence of the
 CC *cag* gene and cytotoxicity suggests that the product of the *cag* gene is
 CC necessary for the transcription, folding, export or function of the
 CC cytotoxin. Alternatively both the *cag* gene and the *ct* gene are absent in
 CC noncytotoxic strains suggesting physical linkage between the genes. The
 CC absence of a typical leader sequence suggests the presence of an
 CC independent export system. Computer searches for promoter regions in the
 CC region upstream from the ATG codon identified sequences resembling either
 CC -10 or -35 regions, however a region with a good consensus to the *E. coli*
 CC or *H. pylori* promoter sequences was not found. The CAI antigen is very
 CC hydrophilic and does not show obvious leader peptide or transmembrane
 CC sequences. The most hydrophilic region is from amino acids 600-900, where
 CC a number of unusual features are observed. There is a repetition of six
 CC sequences EFNGKNQDFSK and EPTIA, and the presence of a stretch of six

CC contiguous asparagines. This protein, and others derived from *H. pylori*,
 CC esp. cytotoxin (CT) or a heat shock protein (see also AA048732 and
 CC AA048734), may be used to treat, prevent and diagnose *H. pylori*
 CC infection. *H. pylori* is the causative agent of type B gastritis, peptic
 CC ulcers and gastric tumours. (Updated on 25-MAR-2003 to correct PW field.)
 CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
 CC correct PA field.)

XX
 XX
 SQ Sequence 5925 BP; 2076 A; 1057 C; 1138 G; 1654 T; 0 U; 0 Other;

Query Match 100.0%; Score 685; DB 2; Length 5925;
 Best Local Similarity 100.0%; Pred. No. 66-159;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AAAAATGGCAAAATAGATTCAGCAAGGTACGCAAGCAAAAGCGACCTTGAAT 60
DB 2782 AAAAATGGCAAAATAGATTCAGCAAGGTACGCAAGCAAAAGCGACCTTGAAT 2841
QY 61 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTAAATGATCATCA 120
DB 2842 TCCGTTAAAGATGTGATCATCAATCAAAAGGTAAAGGTAAATGATCATCA 2901
QY 121 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAGGGTAGAGCAAGGTTAGCCGAT 180
DB 2902 GCGGTATCAGTGGCTAAAGCAAGGGTATTCAGTAGGGTAGAGCAAGGTTAGCCGAT 2961
QY 181 CTCAAAAATTTCTCAAGGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTTCAT 240
DB 2962 CTCAAAAATTTCTCAAGGAGCAATGGCCCAACAGCTCAAAAAATGAAAGTTCAT 3021
QY 241 GCTAGAAAAAATCTGAAATATATCATCCGTTAAGATGTGATGGAACCTTAGTC 300
DB 3022 GCTAGAAAAAATCTGAAATATATCATCCGTTAAGATGTGATGGAACCTTAGTC 3081
QY 301 GGTATGGTATCTCAAGAGAGCCACAACCTTTCTAAAAAATTTTCGACATCAAG 360
DB 3082 GGTATGGTATCTCAAGAGAGCCACAACCTTTCTAAAAAATTTTCGACATCAAG 3141
QY 361 AAAGATTGAATGCAAACTGGAATTTTCAATTAACATTAATGATCTCAAAAC 420
DB 3142 AAAGATTGAATGCAAACTGGAATTTTCAATTAACATTAATGATCTCAAAAC 3201
QY 421 GAACCATTTATGCTAAATTAATAAAGAAAGAGGCAAGCAGCTAGCCCTGAGAA 480
DB 3202 GAACCATTTATGCTAAATTAATAAAGAAAGAGGCAAGCAGCTAGCCCTGAGAA 3261
QY 481 CCCATTTAGCTCAAGTTGCTAAAAAGTAAATGCAAAATTCACCGACTCAATCAATA 540
DB 3262 CCCATTTAGCTCAAGTTGCTAAAAAGTAAATGCAAAATTCACCGACTCAATCAATA 3321
QY 541 GGAATGCTTTGGGTGTTTGTAGGCGCAAGCGGGCTTCCTTTGAAAGGCATGATAA 600
DB 3322 GGAATGCTTTGGGTGTTTGTAGGCGCAAGCGGGCTTCCTTTGAAAGGCATGATAA 3381
QY 601 GTTGATGATCTCAGTAGAGGGCTTTCAGGAATCAGAAATGGCTCAGAAATGAC 660
DB 3382 GTTGATGATCTCAGTAGAGGGCTTTCAGGAATCAGAAATGGCTCAGAAATGAC 3441
QY 661 AATCTCAATCAAGCGGTATCAGAG 685
DB 3442 AATCTCAATCAAGCGGTATCAGAG 3466

```

RESULT 3
 AAT46159
 ID AAT46159 standard; DNA; 19932 BP.

AC AAT46159;
 XX
 DT 30-JUN-1997 (first entry)
 XX
 DE CagI locus.
 XX

```

KW CagI; CagA; virulence factor; exporter molecule; homology; pfl gene;  

KW Bordetella pertussis; VIR B; Agrobacterium tumefaciens; invasion factor;  

KW Salmonella; type I strait; virulence; diagnosis; H. pylori; infection;  

KW vaccine; treatment; duodenal; gastric ulcer; active gastritis;  

KW adenocarcinoma; sg.  

XX  

XX Helicobacter pylori.  

OS  

FH Key Location/Qualifiers  

FT CDS 389..1579  

FT /tag= a  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(1778..2173)  

FT /tag= e  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS 1794..2168  

FT /tag= b  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(2591..3001)  

FT /tag= f  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(3020..3259)  

FT /tag= g  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS 3343..4491  

FT /tag= c  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS 4488..5426  

FT /tag= d  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(5508..6233)  

FT /tag= h  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(6218..6613)  

FT /tag= h  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(6562..7074)  

FT /tag= i  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(7007..7375)  

FT /tag= j  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(7371..7802)  

FT /tag= k  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(7975..8454)  

FT /tag= l  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(8496..8918)  

FT /tag= m  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(8977..9762)  

FT /tag= n  

FT /note= "putative open reading frame; no start codon  

FT given"  

FT CDS complement(9784..10575)  

FT /tag= o  

FT /note= "putative open reading frame; no start codon  

FT given"

```

[illegible]

Qy	121	GGGATATAGGGCTTAAAGCAAGGGTGAATTTTCAGTAGGGTGAAGCAAGCGTTAGCCGAT	180
Db	16909	GGGGATTCAGTGGCTTAAAGCAAGGGTGAATTTTCAGTAGGGTGAAGCAAGCGTTAGCCGAT	16966
Qy	181	CTCAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT	240
Db	16969	CTCAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAATGAAAGTCTCAAT	17028
Qy	241	GGTGAATAAAATCTGAATAATATCATCCGTTAAGATGATGTAATGGAACCGTAGTC	300
Db	17029	GGTGAATAAAATCTGAATAATATCATCCGTTAAGATGTAATGGAACCGTAGTC	17088
Qy	301	GGTATAGGGTATATCTCAAGCAAGCAACACTCTTTCTTAAAACTTTTCGACATCAAG	360
Db	17089	GGTATAGGGTATATCTCAAGCAAGCAACACTCTTTCTTAAAACTTTTCGACATCAAG	17148
Qy	361	AAAGAGTGAATGCAAACTTGGAAATTTCTAATATACATATCAATATGACTCAAAAAC	420
Db	17149	AAAGAGTGAATGCAAACTTGGAAATTTCTAATATACATATCAATATGACTCAAAAAC	17208
Qy	421	GAACCAATTAATGCTAAAGTTAATTAATAAAACAGCGGCAAGCACTAGCTTGAAGAA	480
Db	17209	GAACCAATTAATGCTAAAGTTAATTAATAAAACAGCGGCAAGCACTAGCTTGAAGAA	17268
Qy	481	CCCATTTAGGCTCAAGTTCGCTAAAAAGTTAATGCAAAAAATTGACCGATCATCAATA	540
Db	17269	CCCATTTAGGCTCAAGTTCGCTAAAAAGTTAATGCAAAAAATTGACCGATCATCAATA	17328
Qy	541	GCAAGGGTGGTGGCTGTGTGTATGAGGCAAGCGGGCTTCCCTTTGAAAGGATGATATA	600
Db	17329	GCAAGGGTGGTGGCTGTGTGTATGAGGCAAGCGGGCTTCCCTTTGAAAGGATGATATA	17388
Qy	601	GTTGATGATCTCAGTAGGTAAGTGAAGGCTTTCAAGGATCAAGAAATGGCTCAAGAAATTCAC	660
Db	17389	GTTGATGATCTCAGTAGGTAAGTGAAGGCTTTCAAGGATCAAGAAATGGCTCAAGAAATTCAC	17448
Qy	661	AATCTCAATCAAGCGGTATCAAGAG	685
Db	17449	AATCTCAATCAAGCGGTATCAAGAG	17473
RESULT 4			
ADU05406	1D	ADU05406 standard; DNA; 3558 BP.	
XX	AC	ADU05406;	
XX	XX	27-JAN-2005 (first entry)	
DE	XX	DNA encoding H. pylori antigenic protein HP0547.	
XX	XX	des; gene; antibacterial; antigenic; H. pylori infection.	
OS	XX	Helicobacter pylori.	
PN	XX	MO2004094467-A2.	
PD	XX	04-NOV-2004.	
PF	XX	22-APR-2004; 2004MO-EP004255.	
PR	XX	22-APR-2003; 2003EP-00450097.	
PA	XX	(INTE-) INTERCELL AG.	
PI	XX	Meinke A, Min Bui D, Nagy E, Henics T;	
DR	XX	WPI, 2004-775908/76.	
XX	XX	P-PDB; ADU05584.	
XX	XX	New hyperimmune serum reactive antigens from Helicobacter pylori, and	
XX	XX	encoding nucleic acid molecules, useful for diagnosing, preventing or	

PT treating H. pylori infections.

PS Disclosure; SEQ ID NO 54; 176pp; English

The invention relates to an isolated nucleic acid molecule encoding a hyperimmune serum reactive antigen or its fragment. The composition (including the nucleic acid molecule, hyperimmune serum-reactive antigen or antibody) is useful for manufacturing a medicament or pharmaceutical preparation (e.g. a vaccine) for treating or preventing *H. pylori* infections. The antigen or its fragment may also be used for isolating, purifying and/or identifying an interaction partner of the hyperimmune serum reactive antigen or fragment; for generating a peptide binding to the hyperimmune serum reactive antigen or fragment, where the peptide is selected from anticalins; for manufacturing a functional nucleic acid selected from aptamers and Spiegelmers; or for manufacturing a functional ribonucleic acid selected from ribozymes, antisense nucleic acids and siRNA. The present sequence represents DNA encoding an *H. pylori* antigenic protein.

SQ Sequence 3558 BP; 1335 A; 607 C; 715 G; 901 T; 0 U; 0 Other;

Query Match	Score	DB	Length
77.6%	531.6	13	3558

Matches 581; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

QY	1	AAAAATGGCAAAAAATTAAGGATTTTCAGCAAGTAAACGCAAGCAAAAAAGCACTTGAANAAT	60
Db	2263	AAAAATGGCAAAAAATTAAGGATTTTCAGCAAGGTAATGCAAGCAAAAAAGGACCTTGAAANAAT	2322
QY	61	TCGCTTAAAGATGTGATCTCAATCAAAAGGTAAACGATTAAGTTGATTAATCTCAATCA	120
Db	2323	TCGCTTAAAGATGTGATCTCAATCAAAAGGTAAACGATTAAGTTGATTAATCTCAATCA	2382
QY	121	GCGGTATCAGGGCTAAAGCAACGGGTATTCAGTAGAGTAAAGCAACGGTTAGCCGAT	180
Db	2383	GCGGTATCAGGGCTAAAGCAATGGGCAATTTCAGTAGAGTAAAGCAAGTGTAGCCGAT	2443
QY	181	CTCAAAAATTTCTCAAAAGACCAATTTGGCCCAACAGCTCAAAAATAATGAAGTCTCAAT	240
Db	2443	CTCAAAAATTTCTCAAAAGACCAATTTGGCTCAACAGCTCAAAAATAATGAAGTCTCAAT	2502
QY	241	GCTAGAAAAAATCTGAAATATATCAATCCGTTAAGAAATGGTGTGAATGGAACCTTAATC	300
Db	2503	ACTGAAAAAATTTCTGAACCTAATACCAATCCGTTAAGAAATGGTGTGAATGGAACCTTAATC	2562
QY	301	GGTATAGGGTATCTCAAGCAGAAAGCCCACTCTTTCTAAAACCTTTTGGACATCAAG	360
Db	2563	GGTATAGGGTATCTGGAATTAAGGCCACACGCTCTCGCAAAAATTTTTCGATATCAAG	2622
QY	361	AAAGAGTTGAATGCAAAACCTTGAAATTTCAATTAACAATAACAAT---AATGACCTCAA	417
Db	2623	AAAGATTTGAATGGAATTTAAAAATTTCATTAACAATATTAATGACCTCAAAAACAGC	2682
QY	418	AACGACACCCATTTATGCTAAAGTTAATTAAGAAAGACGGGCAAGCAGCTAGCCTTGA	477
Db	2683	ACGAAACCCATTTATGCTAAAGTTAATTAAGAAAGAAACAGGACAAAGTAGTAGCCCTGA	2742
QY	478	GAAACCCATTTATGCTCAAGTGTCTAAAAAGGTAAATGCAAAAATTGACCGGCTCAATCA	537
Db	2743	GAAACCCATTTATCTCAAGTGTGCTAAAAAGGTAAATGCAAAAATTGACCGGCTCAATCA	2802
QY	538	ATAGCAAGTGTGTTGGTGTGTTGAGGCAACGAGCGGGCTTCCCTTGAAGAAAGCATGAT	597
Db	2803	ATAGCAAGTGTGTTGGGTGTGTTGAGGCAACGAGCGGGCTTCCCTTGAAGAAAGCATGAT	2862
QY	598	AAAGTTGATATCTCAAGTAGTAGGGCTTTCAAGAAATCAAGA	642
Db	2863	AAAGTTGATATCTCAGTAGTAGGGCTTTCAAGTCACTGACCTCGAA	2907

XX ABX66359;
AC
TH

DT 07-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) DNA #958.
XY

KM Protein-protein interaction; ulcer; selected interacting domain; SID,
KW gene; ds.

OS Helicobacter pylori
YY

PN WO200266501-A2.
XX

PD 29-AUG-2002:
XX

28-DEC-2001; 2001WO-EP015428

02-JAN-2001; 2001US-0259302P
XX
XX

PA (INSP) INST PASTEUR

PI Legrain P, Rain J,

DR WPI; 2002-674910/72.

XX

PT useful for identifying

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

The invention describes a complex of protein-protein interactions in *Helicobacter pylori* selected from 421 complexes given in the specification. The complex of protein-protein interactions are useful for screening for agents which modulate the interaction of proteins. Modulating compounds which binds to a targeted bacterial protein may be used for treating or preventing ulcers in a human or animal. This sequence encodes a selected interacting domain (SID), identified via protein-protein interactions

Sequence 1024 BP; 389 A; 179 C; 215 G; 241 T; 0 U; 0 Other;

Query Match	73.4%;	Score 502.6;	DB 6;	Length 1024;
Best Local Similarity	80.6%;	Score 504.3;	DB 4;	Length 1024;

Matches 552; Conservative 0; Mismatches 61; Indels 3; Gaps 1;

QY	30	GGTAAAGCAGCAAAAAGGACCTTGAATAATTCGCTTAAAGATGATCATCAATCAAAA	89
Db	1	GGTAAAGCAGCAAAAAGGACCTTGAATAATTCGCTTAAAGATGATCATCAATCAAAA	60
QY	90	GGTAAAGCAGCAAAAAGGACCTTGAATAATTCGCTTAAAGATGATCATCAATCAAAA	143
Db	61	GGTAAAGCAGCAAAAAGGACCTTGAATAATTCGCTTAAAGATGATCATCAATCAAAA	120
QY	150	TTTCAGTAGGATGAGCAGCGTTAGCGCATCTCAAAAATTTCTCAAGAGCAATTGGC	209
Db	121	TTTCAGTAGGATGAGCAGCGTTAGCGCATCTCAAAAATTTCTCAAGAGCAATTGGC	180
QY	210	CCACACAGTCAAAAAATGAAAGTCTCAATGCTGAAAAAAAATCTGAAATATATCAATC	265
Db	181	TCACACAGTCAAAAAATGAAAGTCTCAATGCTGAAAAAAAATCTGAAATATATCAATC	240
QY	270	CGTTAAGATGCTGTAATGAAACCTTGTGCTGATATGAGTTATCTCAAGCAGAGCCAC	323
Db	241	CGTTAAGATGCTGTAATGAAACCTTGTGCTGATATGAGTTATGAGGCGCAC	300
QY	330	AACCTCTTTCAAAAATTTTGGACATCAAGAAAGAGTTGATGCAAACTTGGAAATTT	389
Db	301	AGCTCTCGCAAAAATTTTGGATATCAAGAAAGATGAAATTTTAAATTTT	360

XX (UYVA-) UNIV VANDERBILT.
 PA Cover TL, Blaser MJ, Tummuuru MKR;
 XX WPI; 1994-151235/18.
 XX P-PSDB; AAR53269.
 DR
 XX
 PT DNA coding tag A gene, from *Helicobacter pylori* - useful for detecting
 PT predispotion to peptic ulceration.
 XX
 PS Claim 27; Page 63-68; 87pp; English.
 XX
 CC Monoclonal antibodies directed against the antigen may be used to detect
 CC tag A antigen presence which is indicative of a predispotion to peptic
 CC ulceration. A ligand e.g. Antibody, specifically reactive with the tag A
 CC antigen can be used to treat peptic ulcers. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 CC
 SQ Sequence 4821 BP; 1722 A; 874 C; 929 G; 1296 T; 0 U; 0 Other;

Query Match 70.8%; Score 485.2; DB 2; Length 4821;
 Best Local Similarity 86.6%; Pred. No. 1.3e-109;
 Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

1 AAAAAATGGCAAAATTAAGGATTTTCAGCAAGGTAACCAAGCAAAAGCACTTGAAT 60
 3319 AAAAAATGGCAAAATTAAGGATTTTCAGCAAGGTAACCAAGCAAAAGCACTTGAAT 3378
 61 TCCGTTAAAGATGTGATCATCATCAAAAGTAACGATTAAGTTGATTAATCTCAATCA 120
 3379 TCCATTAAGATGTGATCATCATCAAAAGTAACGATTAAGTTGATTAATCTCAATCA 3438
 121 GCGGATCACTGCGCTTAAGCAACGGGTGATTTCAAGGAGTAAGCAACGGTTAGCCAT 180
 3439 GCGGATCACTGCGCTTAAGCAACGGGTGATTTCAAGGAGTAAGCAACGGTTAGCCAT 3498
 181 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCAACCAAGCTCAAAAAATTAAGTCTCAAT 240
 3499 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCAACCAAGCTCAAAAAATTAAGTCTCAAT 3558
 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGTGAATGAACCCCTAGTC 300
 3559 GTTGG---AAATCTGAATATATCAATCCGTTAAGATGTGTGAATGAACCCCTAGTC 3615
 301 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTTAAAACTTTTGAGCATCAAG 360
 3616 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTTAAAACTTTTGAGCATCAAG 3675
 361 AAAGGTTGAATGCAAACTTGAATTTCAATTAACATTAATGACTCAAAAC 420
 3676 AAAGGTTGAATGCAAACTTGAATTTCAATTAACATTAATGACTCAAAAC 3735
 421 -----GAACCATTTATGCTAAAGTTAATAAAAAAGAGGCAAGCAAGCTGACCT 474
 3736 GCGGAGAAACCATTTATGCTAAAGTTAATAAAAAAGAGGCAAGCAAGCTGACCT 3795
 475 GAAGAACCATTTATGCTCAAGTTCTAAAGGTAATGCAAAATTTGACCACTCAAT 534
 3796 GAAGAACCATTTATGCTCAAGTTCTAAAGGTAATGCAAAATTTGACCACTCAAT 3855
 535 CAATATGCAAGTGTGGGTGTTGTAAGGCAAGCAAGGCGGCTTCCCTTTGAAAAGCAT 594
 3856 CAAGCAAGCAAGTGTGGGTGTTGTAAGGCAAGGCGGCTTCCCTTTGAAAAGCAT 3915
 595 GATAAGTTGATGATCTCAGTAAGGTAAGGCTTTCAAGGAATCAGAA 642
 3916 GATAAGTTGATGATCTCAGTAAGGTAAGGCTTTCAAGGAATCAGAA 3963

RESULT 8
 AA086728
 ID AA086728 standard; DNA; 4821 BP.

XX
 AC AA086728;
 XX
 DT 25-MAR-2003 (rev15ed)
 DT 29-SEP-1995 (first entry)
 XX
 DE H. pylori taga antigen-encoding DNA.
 XX
 KW Taga; antigen; ulcer; diagnosis; vaccine; da.
 XX
 OS *Helicobacter pylori*.
 XX
 FH Key Location/Qualifiers
 FT CDS 1072..4617
 XX /*tag= a
 XX
 PN US403924-A.
 XX
 PD 04-APR-1995.
 XX
 PF 26-APR-1993; 93US-00053614.
 XX
 PR 13-OCT-1992; 92US-00959940.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Blaser MJ, Tummuuru MKR, Cover TL;
 XX
 DR WPI; 1995-14685/19.
 XX P-PSDB; AAR72593.
 XX

PT New nucleic acid encoding tag A antigen of *Helicobacter pylori* - used to
 PT detect predispotion to peptic ulceration and to produce protein for use
 PT in vaccines, diagnosis etc.
 XX
 PS Disclosure; Col 37-46; 30pp; English.

XX
 CC The full-length sequence of the taga gene of *H. pylori* 84-183 (ATCC
 CC 53726) was obtained from overlapping clones isolated from genomic
 CC libraries. Vectors carrying the gene are used for production of
 CC recombinant taga antigen in host cells. (Updated on 25-MAR-2003 to
 CC correct PR field.)
 XX

SQ Sequence 4821 BP; 1721 A; 874 C; 930 G; 1296 T; 0 U; 0 Other;

Query Match 70.8%; Score 485.2; DB 2; Length 4821;
 Best Local Similarity 86.6%; Pred. No. 1.3e-109;
 Matches 561; Conservative 0; Mismatches 78; Indels 9; Gaps 2;

1 AAAAAATGGCAAAATTAAGGATTTTCAGCAAGGTAACCAAGCAAAAGCACTTGAAT 60
 3319 AAAAAATGGCAAAATTAAGGATTTTCAGCAAGGTAACCAAGCAAAAGCACTTGAAT 3378
 61 TCCGTTAAAGATGTGATCATCATCAAAAGTAACGATTAAGTTGATTAATCTCAATCA 120
 3379 TCCATTAAGATGTGATCATCATCAAAAGTAACGATTAAGTTGATTAATCTCAATCA 3438
 121 GCGGATCAAGGCTTAAGCAACGGGTGATTTCAAGTGGGTGAAGCAACGTTAGCCAT 180
 3439 GCGGATCAAGGCTTAAGCAACGGGTGATTTCAAGTGGGTGAAGCAACGTTAGCCAT 3498
 181 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCAACCAAGCTCAAAAAATTAAGTCTCAAT 240
 3499 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCAACCAAGCTCAAAAAATTAAGTCTCAAT 3558
 241 GCTGAAAAAATCTGAATATATCAATCCGTTAAGATGTGTGAATGAAACCTTAGTC 300
 3559 GTTGG---AAATCTGAATATATCAATCCGTTAAGATGTGTGAATGAAACCTTAGTC 3615
 301 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTTAAAACTTTTGAGCATCAAG 360
 3616 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTTAAAACTTTTGAGCATCAAG 3675

	RESULT	12
	AAT67410	
ID	AAT67410	standard; DNA; 3534 bp.
AC	AAT67410;	
CC		
DT	08-JUN-1997	(first entry)
DE	H. pylori cytoplasmic protein ORF 16459375..aa.	
DD	Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.	
XX	Helicobacter pylori.	
XX		
OS		
PPH	Key	Location/Qualifiers
PFT	CDS	1..3534
PFT		/*tag= a
XXX		/note= "no stop codon is given"
PPN	M09640893-AI.	
XX		
PD	19-DEC-1996.	
PP		
XX	06-JUN-1996;	96MO-US009122.
PR	07-JUN-1995;	95US-00487032.
OR	01-APR-1996;	96US-00630405.

The present sequence encodes a Helicobacter pylori cytoplasmic protein.. The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors. The genomic sequence of H. pylori (ATCC 56679) was determined from overlapping sequences generated by mechanically shearing the bacterial DNA. The contigs generated were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Db 2782 GCAAGTGGTTGGTGATGTAGGCGAAGCAGCGCTCTCTTTGAAGAGCATGATMAA 2841
 QY 601 GTTGATGATCTCAGTAAGGAGGCTTTCAAGGATCAAGAA 642
 Db 2842 GTTGATGATCTCAGTAAGGAGGCTTTCAAGCTACCAAGAA 2883

RESULT 13
 AAT68137
 ID AAT68137 standard; DNA; 3549 BP.

AC AAT68137;

DT 17-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein ORF 13ee12016orf74.

KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis; ds.

OS Helicobacter pylori.

Key Location/Qualifiers

FT CDS 1..3549

FT /note="no stop codon given"

PN W09640893-A1.

PD 19-DEC-1996.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.

PR 01-APR-1996; 96US-00630405.

XX (ASTR) ASTRA AB.

PI Smith D, Berglindh OT, Mellgaerd BL;

DR WPI, 1997-052306/05.

XX P-PSDB; AAM20884.

PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 useful for vaccines to treat or prevent H. pylori infection, and to
 detect Helicobacter.

PS Claim 9; Page 921-22; 1481pp; English.

XX This sequence encodes a H. pylori cytoplasmic protein. The protein may be
 used in a vaccine to prevent or treat H. pylori infection or to identify
 H. pylori polypeptide binding compounds, useful as potential H. pylori
 life cycle activators or inhibitors. The genomic sequence of H. pylori
 (ATCC 55679) was determined from overlapping contigs generated by
 mechanically shearing the bacterial DNA. The sequences were analysed for
 ORF of at least 180 nucleotides, and the predicted coding regions defined
 by computer evaluation. To identify likely H. pylori antigens for vaccine
 development, the amino acid sequences predicted from various ORF were
 analysed for significant homology to other known or exported membrane
 proteins. Having identified and determined the sequences of interest,
 particular regions can be isolated from H. pylori by PCR amplification
 for recombinant polypeptide production, e.g. in E. coli hosts

CC Sequence 3549 BP; 1312 A; 604 C; 723 G; 910 T; 0 U; 0 Other;

Query Match 69.7%; Score 477.2; DB 2; Length 3549;

Best Local Similarity 86.9%; Pred. No. 1.1e-107;

Matches 558; Conservative 0; Mismatches 33; Indels 51; Gaps 1;

QY 1 AAAATGCAAAAATTAAGGATTTCAAGCAAGTAAGCAAGCAAAAAGCACTTGAAAT 60

Db 2308 AAAATGCAAAAATTAAGGATTTCAAGCAAGTAAGCAAGCAAAAAGCACTTGAAAT 2367

QY 61 TCCGTTAAAGATGTATCATCAATCAAAAAGTAACGATTAAGTATATCATCA 120

Db 2368 TCCATTTAAAGATGTATCATCAATCAAAAAGTAACGATTAAGTATATCATCA 2427

QY 121 GCGGTATCAGTGGCTAAAGCAAGGGGTGATTTCAATAGGTTAGCAACGCTTACCGAT 180

Db 2428 GCGGTATCAGTGGCTAAAGCAAGGGGTGATTTCAATAGGTTAGCAACGCTTACCGAT 2487

QY 181 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCTCAAT 240

Db 2488 CTCAAAAATTTCTCAAGAGCAATGGCCCAAGCTCAAAAAATGAAGTCTCAAT 2547

QY 241 GCTAGAAAAAATCTGAATAATATATATATCCGTTAAGAAATGTGAATGAAACCTTAATC 300

Db 2548 ACTGCAAAAAATTTCTGCACTATACCAATCCGTTAAGAAATGTGAATGAAACCTTAATC 2607

QY 301 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTCTAATAAACTTTTGGACATCAAG 360

Db 2608 GGTAAATGGTTATCTCAAGCAAGCAACCACTCTTTCTAATAAACTTTTGGACATCAAG 2667

QY 361 AAAGAGTTGAATGCAAAAATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420

Db 2668 AAAGAGTTGAATGCAAAAATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2727

QY 421 GAACCAATTTATGCTAAAGTTAATTAATAAAGAAAGCAAGGCAAGCTAGCCTTGAAGAA 480

Db 2728 AGCAC-----AGAA 2736

QY 481 CCCATTTACGCTCAAGTCTTAAAGGTAATGCAAAAATTTGACCGACTCAATCAATA 540

Db 2737 CCCATTTACGCTCAAGTCTTAAAGGTAATGCAAAAATTTGACCGACTCAATCAATA 2796

QY 541 GCAAGTGGTTGGGTGTGAGGGGCAAGCGGGCTCCCTTGAAGAAGCATGATAA 600

Db 2797 GCAAGTGGTTGGGTGTGAGGGGCAAGCGGGCTCCCTTGAAGAAGCATGATAA 2856

QY 601 GTTGATGATCTCAGTAAGGAGGCTTTCAAGGATCAAGAA 642

Db 2857 GTTGATGATCTCAGTAAGGAGGCTTTCAAGGATCAAGAA 2898

RESULT 14

ABX66128
 ID ABX66128 standard; DNA; 1024 BP.

XX ABX66128;

DT 07-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) DNA #727.

XX Protein-protein interaction; ulcer; selected interacting domain; SID;

KW gene; ds.

OS Helicobacter pylori.

XX W020026501-A2.

PN 29-AUG-2002.

PD 28-DEC-2001; 2001WO-BP015428.

PP 02-JAN-2001; 2001US-0259302P.

XX (HYBR-) HYBRIGENICS.

PA (INSP) INST PASTEUR.

PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

DR WPI, 2002-674910/72.

XX P-PSDB; AB051384.

Tue Mar 7 12:58:25 2006

Search completed: March 6, 2006, 22:05:53
Job time : 558 secs

us-09-360-685c-26.png

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd

Run on: March 6, 2006, 21:27:13 ; Search time 3660 Seconds

10638.733 Million cell updates/sec

Title: US-09-360-685C-26

Sequence: 1 aaatgcaaaataaga.....caatcaagcgtatcagaag 685

Scoring table: IDENTITY_NUC

Searched: 5883141 beqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum	DB	seq	length:	0
Maximum	DB	seq	length:	2000000000

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  gb_da:*
2:  gb_in:*
3:  gb_env:*
4:  gb_cm:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pr:*
9:  gb_ro:*
10: gb_sts:
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_htg:*
15: gb_pl:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	685	100.0	5925	1	HPCAT	X70039 H.pylori ca
2	685	100.0	5925	6	A76409	A76409 Sequence 4
3	685	100.0	5925	6	AR099343	AR099343 Sequence
4	685	100.0	5925	6	AR112380	AR112380 Sequence
5	685	100.0	5925	6	AX030324	AX030324 Sequence
6	685	100.0	5925	6	BD000835	BD000835 Helicobac
7	685	100.0	5925	6	BD000839	BD000839 Helicobac
8	685	100.0	20229	1	AF282853	AF282853 Helicobac
9	683.4	99.8	3444	1	AB015416	AB015416 Helicobac
10	601.6	87.9	3426	1	DQ011619	DQ011619 Helicobac
11	597	87.2	978	1	AB110963	AB110963 Helicobac
12	595.6	86.9	3693	1	DQ067454	DQ067454 Helicobac
13	589.2	86.0	26663	1	AY330638S2	AY330638 Helicobac
14	584.4	85.3	3543	1	AB090088	AB090088 Helicobac
15	584.4	85.3	26771	1	AY330637	AY330637 Helicobac
16	584.4	85.3	37037	1	AB120425	AB120425 Helicobac
17	579.4	84.6	981	1	AB090151	AB090151 Helicobac
18	579.4	84.6	3447	1	AB190949	AB190949 Helicobac

19	570	83.2	3540	1	AB015414	Helicobac
20	566.6	82.5	815	6	AR679176	Sequence
21	565.2	82.5	3540	1	AF247651	Helicobac
22	565	82.5	801	1	AF479032	Helicobac
23	558.4	81.5	3471	1	AF884088	Helicobac
24	554	80.9	3744	1	AF202973	Helicobac
25	553.4	80.6	3744	1	AB015413	Helicobac
26	549.2	80.2	3744	1	AF000337	Helicobac
27	549.2	80.2	3744	1	AF001357	Helicobac
28	547.6	79.9	7498	1	AF330640S3	Helicobac
29	547.6	79.9	26276	1	AFY33064S2	Helicobac
30	544.2	79.4	1323	1	AF222809	Helicobac
31	541.2	79.0	4093	1	AF0683352	Helicobac
32	539.6	78.8	1077	1	AB090413	Helicobac
33	539.6	78.8	3552	1	AB109640	Helicobac
34	534.8	78.1	933	1	AFY571191	Helicobac
35	534.8	78.1	933	1	AFY571192	Helicobac
36	534.8	78.1	933	1	AFY571193	Helicobac
37	534.8	78.1	933	1	AFY571194	Helicobac
38	534.8	78.1	933	1	AFY571195	Helicobac
39	534.8	78.1	933	1	AFY571196	Helicobac
40	534.4	78.0	1224	1	AF222808	Helicobac
41	532.4	77.7	3558	1	DQ011620	Helicobac
42	531.6	77.6	3558	6	CQ090322	Sequence
43	531.6	77.6	3651	1	AB090082	Helicobac
44	531.6	77.6	11141	1	AB090059	Helicobac
45	531.6	77.6	37459	1	AB104020	Helicobac

ALIGNMENTS

RESULT 1	
HPCAI	5925 bp DNA linear BCT 22-JUL-1993
LOCUS	
DEFINITION	H.pylori cal gene for cytotoxicity associated immunodominant antigen.
ACCESSION	X70039
VERSION	X70039.1 GI:394912
KEYWORDS	Cytotoxicity associated immunodominant antigen.
SOURCE	Helicobacter pylori
ORGANISM	Helicobacter pylori Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Helicobacteraceae; Helicobacter.
REFERENCE	1
AUTHORS	Covacci,A., Censini,S., Bugnoli,M., Petracca,R., Burroni,D., Machia,G., Massone,A., Papili,E., Xiang,Z., Figura,N. and Rappunli,R. Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer Proc. Natl. Acad. Sci. U.S.A. 90 (12), 5791-5795 (1993)
JOURNAL	8516329
PUBMED	2 (bases 1 to 5925)
REFERENCE	Covacci,A.
AUTHORS	Direct Submission
TITLE	Submitted (12-JAN-1993) A. Covacci, Iris, Immunobiological Research
JOURNAL	Inst Siena, Via Fiorentina, 1, 53100 Siena, ITALY
FEATURES	Location/Qualifiers
Source	1..5925
	/organism="Helicobacter pylori"
	/mol_type="genomic DNA"
	/strain="CCUG 17874"
	/db_xref="taxon:210"
	/clone="57/D_64/4_007_24_A1_G5_A17_B1"
	/clone_id="lambda gt11 AR12/plasmid pHDIIIIT"
	1..534
promoter	prim_transcript 324..4032
	/note="transcript B"
prim_transcript	430..4032
	/note="transcript A"
RBS	525..529
gene	535..5978
	/gene="cal"

Dd		3262	CCCAATTACGCTCAAGTTGGCTATAAAAGTAAATGCAAAAATTAGACCACACTCAATCAATAA	3322
Oy		541	GCAAGTGTTTTGGGTGGTGTGTAGCGCAGACGACGGGGCTTCCTTTGA AAAAGCGATGATAAA	600
Dd		3322	GCAAGTGTTTTGGGTGGTGTGTAGCGCAGACGACGGGGCTTCCTTTGA AAAAGCGATGATAAA	3381
Oy		601	GTTCATGATCTCAGTAAGGTAGGGCTTTCAAAGAAATCAAGAAATTTGGCTCAGAAAATTGAC	660
Dd		3382	GTTCATGATCTCAGTAAGGTAGGGCTTTCAAAGAAATCAAGAAATTTGGCTCAGAAAATTGAC	3441
Oy		661	AATCTCAATCAAGCGGTATCAGAAG	685
Dd		3442	AATCTCAATCAAGCGGTATCAGAAG	3466
 RESULT 3 AR099343				
LOCUS	AR099343	5925 bp	DNA	linear
DEFINITION	Sequence 4 from patent US 6077706.			PAT 14-FEB-2001
ACCESSION	AR099343			
VERSION	AR099343.1	GI:12809109		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 5925)			
TITLE	Covacci,A., Bagnoli,M., Telford,J., Macchia,G. and Rappuoli,R.			
JOURNAL	Helicobacter pylori proteins useful for vaccines and diagnostics			
FEATURES	Patent: US 6077706-A 4 20-JUN-2000;			
source	Location/Qualifiers			
	1..5925			
	/organism="unknown"			
	/mol_type="unassigned DNA"			

Db	Accession	Sequence	Length	Source	Organism	Reference	Title	Journal	Features
Db	3262	CCCATTTACGCTCAAAGTTGCTAAAAAGGTAATGCAAAATTTGACCCGATCAATCAATA	3321						
Qy	541	GCAAGTGTGTTGGGTGTGTGTGTAAGGGCAAGCAGCGGGCTTCCTTTGAAAAAGGCATGATAA	600						
Db	3322	GCAAGTGTGTTGGGTGTGTGTGTAAGGGCAAGCAGCGGGCTTCCTTTGAAAAAGGCATGATAA	3381						
Qy	601	GTTGATGATCTCAAGTAAGTAAGGGCTTTCAAGAAATCAAGATTTGGCTCAGAAAAATTGAC	660						
Db	3382	GTTGATGATCTCAAGTAAGTAAGGGCTTTCAAGAAATCAAGATTTGGCTCAGAAAAATTGAC	3441						
Qy	661	AATCCATCAATCAAGCGGTATCAAGG	685						
Db	3442	AATCTCAATCAAGCGGTATCAAGG	3466						
RESULT 4									
LOCUS	AR112380	5925 bp	DNA	linear	PAT 16-MAY-2001				
DEFINITION	Sequence 4 from patent US 6130059.								
ACCESSION	AR112380								
VERSION	AR112380.1	GI:14092280							
KEYWORDS									
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	Unclassified.								
AUTHORS	1 (bases 1 to 5925)								
TITLE	Covacci, A., Bugnoli, M., Telford, J., Macchia, G. and Rappuoli, R.								
JOURNAL	Helicobacter pylori proteins useful for vaccines and diagnostics								
FEATURES	Parent: US 6130059-A 4 10-Oct-2000;								
	Location/Qualifiers								
	1..5925								
	/organism="unknown"								
	/mol_type="unassigned DNA"								

Db	3262	CCCAATTACGCTCAAGTTGCTCTAAAAAGTAAATGAAAAATTTGACCCACTCAATCAATA	3322
Qy	541	GCAAGTGGTTTGGGTGTTGTAGGGCAGACGCGGGCTTCCCTTTGAAAAGCGATGATAA	600
Db	3322	GCAAGTGGTTTGGGTGTTGTAGGGCAGACGCGGGCTTCCCTTTGAAAAGCGATGATAA	3381
Qy	601	GTTGATGATCTCAGTAGGTAGGGCTTTCAGAGAAATCAAGAAATTTGGCTCAGAAATTGAC	660
Db	3382	GTTGATGATCTCAGTAGGTAGGGCTTTCAGAGAAATCAAGAAATTTGGCTCAGAAATTGAC	3441
Qy	661	AATCTCAATCAAGCGGTATCAGAG	685
Db	3442	AATCTCAATCAAGCGGTATCAGAG	3466
RESULT 3			
LOCUS	AR099343	5925 bp	DNA
DEFINITION	Sequence 4 from patent US 6077706.	linear	PAT 14-FEB-2001
ACCESSION	AR099343		
VERSION	AR099343.1	GI:12809109	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5925)		
AUTHORS	Covacci, A., Bagnoli, M., Telford, J., Macchia, G. and Rappuoli, R.		
TITLE	Helicobacter pylori proteins useful for vaccines and diagnostics		
JOURNAL	Patent: US 6077706-A 4 20-JUN-2000;		
FEATURES	Location/Qualifiers		
source	1..5925		
	/organism="Unknown"		
	/mol_type="unassigned DNA"		

Db	Accession	Source	Organism	Reference	Location/Qualifiers	mol_type
Db	3262	CCCAATTACGCTCAAAGTTGCTAAAAAGGTAATGCAAAATTTGACCCGATCAATCAATA	3321			
Qy	541	GCAAGTGTGTTGGGTGTGTGTAGGGCAAGCAGCGGGCTTCCTTTGAAAAAGCATGATAA	600			
Db	3322	GCAAGTGTGTTGGGTGTGTGTAGGGCAAGCAGCGGGCTTCCTTTGAAAAAGCATGATAA	3381			
Qy	601	GTTGATGATCTCAAGTAAAGTAGGGCTTCAAGAAATCAAGATTTGGCTCAGAAAAATTGAC	660			
Db	3382	GTTGATGATCTCAAGTAAAGTAGGGCTTCAAGAAATCAAGATTTGGCTCAGAAAAATTGAC	3441			
Qy	661	AATCCATCAATCAAGCGGTATCAGAG	685			
Db	3442	AATCTCAATCAAGCGGTATCAGAG	3466			
RESULT 4						
AR112380						
LOCUS	AR112380	5925 bp	DNA	linear	PAT 16-MAY-2001	
DEFINITION	Sequence 4 from patent US 6130059.					
ACCESSION	AR112380					
VERSION	AR112380.1	GI:14092280				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 5925)					
TITLE	Covacci, A., Bugnoli, M., Telford, J., Macchia, G. and Rappuoli, R.					
JOURNAL	Helicobacter pylori proteins useful for vaccines and diagnostics					
FEATURES	Patent: US 6130059-A 4 10-Oct-2000;					
Source	Location/Qualifiers					
	1..5925					
	/organism="unknown"					
	/mol_type="unassigned DNA"					

Db 3262 CCATTTAGGCTCAAGTTGCTAAAGTAAATGCAAAATTTGACCACTCAATCAATA 3321
 QY 541 GCAAGTGGTTGGTGTGTGAGGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 600
 Db 3322 GCAAGTGGTTGGTGTGTGAGGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 3381
 QY 601 GTTGATGATCTCAGTAAGAGGCTTTCAGAGGATCAAGATTGGCTCGAAATTTGAC 660
 Db 3382 GTTGATGATCTCAGTAAGAGGCTTTCAGAGGATCAAGATTGGCTCGAAATTTGAC 3441
 QY 661 AATCTCAATCAAGCGGTATCAAG 685
 Db 3442 AATCTCAATCAAGCGGTATCAAG 3466

RESULT 5
 AX030324 5925 bp DNA linear PAT 16-SEP-2000
 LOCUS Sequence 3 from Patent EP0967279.
 ACCESSION AX030324
 VERSION AX030324.1 GI:10190490
 KEYWORDS
 SOURCE Helicobacter pylori
 ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 Helicobacteriaceae; Helicobacter.

REFERENCE
 AUTHORS Covacci, A., Rappelli, R., Bugnoli, M., Telford, J., Macchia, G.
 TITLE Helicobacter pylori cytotoxin useful for vaccines and diagnostics
 JOURNAL Patent: EP 0967279-A 3 29-DEC-1999;
 CHIRON SPA (IT)

FEATURES
 source 1..5925
 /organism="Helicobacter pylori"
 /mol_type="unassigned DNA"
 /db_xref="taxon:210"

ORIGIN

Query Match 100.0%; Score 685; DB 6; Length 5925;
 Best Local Similarity 100.0%; Pred. No. 5, 2e-118;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 60
 Db 2782 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 2841
 QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 120
 Db 2842 TCCGTTAAAGATGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 2901
 QY 121 GCGGTATCAAGTGGCTAAAGCAAGGCTATTCAGTAGGTTAGCAAGGTTAGCCAT 180
 Db 2902 GCGGTATCAAGTGGCTAAAGCAAGGCTATTCAGTAGGTTAGCAAGGTTAGCCAT 2961
 QY 181 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCCCAAGCTCAAAATTAAGAGTTCAT 240
 Db 2962 CTCAAAAATTTCTCAAAAGAGCAATTTGCCCCCAAGCTCAAAATTAAGAGTTCAT 3021
 QY 241 GCTAGAAAAAATCTGAAT 300
 Db 3022 GCTAGAAAAAATCTGAAT 3081
 QY 301 GGTATGGGTTATCTCAAGCAAGAGCAAGCTTTCTTAAATCTTTGCGACATCAAG 360
 Db 3082 GGTATGGGTTATCTCAAGCAAGAGCAAGCTTTCTTAAATCTTTGCGACATCAAG 3141
 QY 361 AAAGGTTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGATCACTAAAG 420
 Db 3142 AAAGGTTGAATGCAAACTTGAATTTCAATTAACATTAACATTAATGATCACTAAAG 3201
 QY 421 GAACCAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
 Db 3202 GAACCAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3261

QY 481 CCATTTAGGCTCAAGTTGCTAAAGTAAATTCGAAATTTGACCACTCAATCAATA 540
 Db 3262 CCATTTAGGCTCAAGTTGCTAAAGTAAATTCGAAATTTGACCACTCAATCAATA 3321
 QY 541 GCAAGTGGTTGGTGTGTGAGGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 600
 Db 3322 GCAAGTGGTTGGTGTGTGAGGGCAAGCAGCGGCTTCCCTTTGAAAAGGCAATGATAA 3381
 QY 601 GTTGATGATCTCAGTAAGAGGCTTTCAGAGGATCAAGATTGGCTCGAAATTTGAC 660
 Db 3382 GTTGATGATCTCAGTAAGAGGCTTTCAGAGGATCAAGATTGGCTCGAAATTTGAC 3441
 QY 661 AATCTCAATCAAGCGGTATCAAG 685
 Db 3442 AATCTCAATCAAGCGGTATCAAG 3466

RESULT 6
 BD000835 5925 bp DNA linear PAT 31-JAN-2002
 LOCUS Helicobacter pylori proteins useful for vaccines and diagnostics.
 ACCESSION BD000835
 VERSION BD000835.1 GI:18625394
 KEYWORDS JP 2000350591-A/3.
 SOURCE unidentified
 ORGANISM unidentified

REFERENCE
 AUTHORS Covacci, A., Bugnoli, M., Telford, J., Macchia, G., and Rappelli, R.
 TITLE Helicobacter pylori proteins useful for vaccines and diagnostics
 JOURNAL Patent: JP 2000350591-A 3 19-DEC-2000;
 CHIRON SPA

COMMENT

OS Unidentified
 PN JP 2000350591-A/3
 PD 19-DEC-2000
 PF 26-APR-2000 JP 2000126695
 PR 02-MAR-1992 IT FI92A00052, 25-JAN-1993 GB PCTEP9300158 PI
 ANTONELLO COVACCI, MASSIMO BUGNOLI, JOHN TELFORD, PI GIOVANNI
 MACCHIA,
 PI RINO RAPPUOLI
 PC C12N15/09, A61K38/00, A61K39/106, A61K39/39, A61K48/00, A61P1/04,
 PC A61P31/04,
 PC A61P35/00, C07K14/195, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
 C12P21/02,
 PC C12O1/68, G01N33/569, C12N15/09, C12R1/01, C12N1/21, C12R1/19,
 PC C12P21/02, C12R1/19, C12N15/00, A61K37/02, C12N5/00, C12N15/00,
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key
 FT source 1..5925
 /location/Qualifiers
 1..5925
 /organism="Unidentified",
 /mol_type="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

FEATURES

source 1..5925
 /organism="Unidentified",
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"

ORIGIN

Query Match 100.0%; Score 685; DB 6; Length 5925;
 Best Local Similarity 100.0%; Pred. No. 5, 2e-118;
 Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 60
 Db 2782 AAAAAAGCAAAATTAAGATTTTCAGCAAGTAAAGCAAGCAAAAGGACCTTGAAT 2841
 QY 61 TCCGTTAAAGATGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 120
 Db 2842 TCCGTTAAAGATGATCATCAATCAAAAGGTAAACGATTAAGTTGATTCATCA 2901
 QY 121 GCGGTATCAAGTGGCTAAAGCAAGGCTATTCAGTAGGTTAGCAAGGTTAGCCAT 180

```

Db      2902 GCGGATCACTGCTAAAGCAACGGGTGATTCAGTAGGGTAGAGCAAGCGTTAGCGAT 2961
Qy      181 CTCAAAAATTTCTCAAAAGGCAATGGCCCAACAAGCTCAAAAAATGAAATCTCAAT 240
Db      2962 CTCAAAAATTTCTCAAAAGGCAATGGCCCAACAAGCTCAAAAAATGAAATCTCAAT 3021
Qy      241 GCTAGAAAAAATCTGAAATATATATCAATCCGTTAAGATGATGTAATGGAACCTAGTC 300
Db      3022 GCTAGAAAAAATCTGAAATATATATCAATCCGTTAAGATGATGTAATGGAACCTAGTC 3081
Qy      301 GGTAAATGGTTATCTCAAGCAAGACCAACTCTTTCTAAAACTTTTGGACATCAAG 360
Db      3082 GGTAAATGGTTATCTCAAGCAAGACCAACTCTTTCTAAAACTTTTGGACATCAAG 3141
Qy      361 AAGAGGTGAATGCAAAAATCTGAAATTTCAATACATACATATATGATCTCAAAAC 420
Db      3142 AAGAGGTGAATGCAAAAATCTGAAATTTCAATACATACATATATGATCTCAAAAC 3201
Qy      421 GAACCAATTTATCTAAAGTTAATAAAGAAAGCAAGGCAAGCAGTACCTTGAAGA 480
Db      3202 GAACCAATTTATCTAAAGTTAATAAAGAAAGCAAGGCAAGCAGTACCTTGAAGA 3261
Qy      481 CCCATTACGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACGATCAATCAAT 540
Db      3262 CCCATTACGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACGATCAATCAAT 3321
Qy      541 GCAAGGTGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 600
Db      3322 GCAAGGTGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 3381
Qy      601 GTTGATGATCTCAGTAGGAGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAATTTGAC 660
Db      3382 GTTGATGATCTCAGTAGGAGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAATTTGAC 3441
Qy      661 AATTCATCAACGCGGTATCAGAG 685
Db      3442 AATTCATCAACGCGGTATCAGAG 3466

RESULT 7
BD000839 5925 bp DNA linear PAT 31-JAN-2002
LOCUS Helicobacter pylori proteins useful for vaccines and diagnostics.
DEFINITION BD000839
ACCESSION BD000839
VERSION BD000839.1 GI:18625398
KEYWORDS JP 2000333686-A/3.
SOURCE unidentifed
ORGANISM unidentifed.

REFERENCE
1 (bases 1 to 5925)
AUTHORS Covacci,A., Bugnoli,M., Telford,J., Macchia,G. and Rappuoli,R.
TITLES Helicobacter pylori proteins useful for vaccines and diagnostics
JOURNAL Patent: JP 2000333686-A 3 05-DEC-2000;
CHIRON SPA

COMMENT
OS Unidentifed
PN JP 2000333686-A/3
PD 05-DEC-2000
PF 26-APR-2000 JP 2000126696
PR 26-MAR-1992 IT FI92A00052.25-JAN-1993 GB PCTEP9300158 PI
ANTONELLO COVACCI,MASSIMO BUGNOLI,JOHN TELFORD, PI GIOVANNI
MACCHIA.
PI RINO RAPPUOLI
PC C12N15/09,A61K38/00,A61K39/106,A61P1/04,C07K14/195, PC
C12N9/80//C12P21/02,
PC (C12N15/09,C12R1:01),C12N15/00,A61K37/02,(C12N15/00,C12R1:01)
CC Strandedness: Single;
CC Topology: Linear;
CC Key Location/Qualifiers
FH source 1..5925
FT Location/Qualifiers
1..5925
FEATURES
source Location/Qualifiers
1..5925

```

```

/organism="unidentifed"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 100.0%; Score 685; DB 6; Length 5925;
Best Local Similarity 100.0%; Pred. No. 5.2e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AAAAAATGCAAAAATTAAGGATTTCAAGCAAGTACGCAAGCAAAAAGCGACCTTGAAAT 60
Db      2782 AAAAAATGCAAAAATTAAGGATTTCAAGCAAGTACGCAAGCAAAAAGCGACCTTGAAAT 2841
Qy      61 TCCGTTAAAGATGTATCTCAATCAAAAAGTAAAGGATTAAGTTGATATCTCAATCA 120
Db      2842 TCCGTTAAAGATGTATCTCAATCAAAAAGTAAAGGATTAAGTTGATATCTCAATCA 2901
Qy      121 GCGGTACAGTGGCTAAAGCAACGGGTGATTTCAATAGGGGTAGACCAAGCTTAGCCGAT 180
Db      2902 GCGGTACAGTGGCTAAAGCAACGGGTGATTTCAATAGGGGTAGACCAAGCTTAGCCGAT 2961
Qy      181 CTCAAAAATTTCTCAAGGAGCAATTTGCCCACAAGCTCAAAAAATGAAAGTCTCAAT 240
Db      2962 CTCAAAAATTTCTCAAGGAGCAATTTGCCCACAAGCTCAAAAAATGAAAGTCTCAAT 3021
Qy      241 GCTAGAAAAAATCTGAAATATATCAATCCGTTAAGATGATGTAATGGAACCTAGTC 300
Db      3022 GCTAGAAAAAATCTGAAATATATCAATCCGTTAAGATGATGTAATGGAACCTAGTC 3081
Qy      301 GGTAAATGGTTATCTCAAGCAAGCACAACCTCTTTCTAAAACTTTTGGACATCAAG 360
Db      3082 GGTAAATGGTTATCTCAAGCAAGCACAACCTCTTTCTAAAACTTTTGGACATCAAG 3141
Qy      361 AAGAGGTGAATGCAAAAATCTGAAATTTCAATACATACATATATGACTCAAAAC 420
Db      3142 AAGAGGTGAATGCAAAAATCTGAAATTTCAATACATACATATATGACTCAAAAC 3201
Qy      421 GAACCAATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAGTAGCTTGAAGA 480
Db      3202 GAACCAATTTATGCTAAAGTTAATAAAGAAAGCAAGGCAAGCAGTAGCTTGAAGA 3261
Qy      481 CCCATTACGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACGATCAATCAAT 540
Db      3262 CCCATTACGCTCAAGTGTCTAAAGAAAGTAAATGCAAAATTTGACGATCAATCAAT 3321
Qy      541 GCAAGGTGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 600
Db      3322 GCAAGGTGTTGGGTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 3381
Qy      601 GTTGATGATCTCAGTAGGAGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAATTTGAC 660
Db      3382 GTTGATGATCTCAGTAGGAGGCTTTCAAGGAATCAAGAAATTTGCTCAGAAATTTGAC 3441
Qy      661 AATTCATCAACGCGGTATCAGAG 685
Db      3442 AATTCATCAACGCGGTATCAGAG 3466

RESULT 8
AF282853 20229 bp DNA linear BCT 16-AUG-2000
LOCUS Helicobacter pylori cagI pathogenicity island, complete sequence.
DEFINITION AF282853 U60176
ACCESSION AF282853
VERSION AF282853.1 GI:8843984
KEYWORDS
SOURCE
ORGANISM Helicobacter pylori
Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 (bases 1 to 20229)
AUTHORS Censini,S., Lange,C., Xiang,Z., Crabtree,J.E., Ghiara,P.,
Borodovsky,M., Rappuoli,R. and Covacci,A.
TITLES cagI, a pathogenicity island of Helicobacter pylori, encodes type

```

[illegible]

```

RBS      /gene="cagm"
          3628..3631
          /gene="cagm"
          /note="putative"
          3647..4777
CDS      /gene="cagm"
          /note="HP537; cag16; JHP485"
          /codon_start=1
          /transl_table=11
          /product="Cagm"
          /protein_id="AAC44695.1"
          /db_xref="GI:1752689"
          /translation="MLAKIVFSSILVAFGLSANVEQFSPFNEIKKEQEVAKEDAL
          KARKKILNTHDFLEDLIPRKOKIKELMDHRAVYLSDLKVKKKKKEKLEKREGLT
          TAKSKAYGDELEAKONPLKYLKLPNPYAVVLYNOEPTTKEDKRLSYTPQVTSIF
          KKTATATTKDRAQLLQMGVSLDEBQKASRLALSFKOAIIEYSNNISMLSRKELD
          NIDYVLQLENNKFSKAKDIAQKATWLLFNSERLAFSMAIDIKENKYLGYAPENL
          LKNVKDVELNTLTKNFTNOKSFAQOKKCLVLDSFNFDTQSKKSLIKKTNEVNI
          VDSDDPMMSDKTVMQKHYKLPFFKTVSAYRNNAKNPF"
          4781..5712
          /gene="cagm"
          4781..4785
          /gene="cagm"
          /note="putative"
          4792..5712
          /gene="cagm"
          /note="HP538; cag17; JHP486"
          /codon_start=1
          /transl_table=11
          /product="Cagm"
          /protein_id="AAF80207.2"
          /db_xref="GI:9832971"
          /translation="MKSIRPKKLSVALYSLYVYGALNLTALLPEYIKELVALGFK
          IKTLVQRHDKETLKSEKFPATNALREKLNDRARLEQIKNIIEAFKKNSSVQKKA
          AKRKLQELNEINANPLNDPNNGNSSTETKSNDSNDEKINVSFVCPAPLVPD
          EWRTPBIBIINECIISNDYDLKRLKIDIKDKILAPLEIKOIEITENNKFSRO
          HLSGLKLTNNNSNRTFLASCAICERKKEMQENNYQDTTNAISFGTTDPKEMNAK
          DTAFSNRSKSELPNSVINQIROSIAHGKK"
          complement(5794..6507)
          /gene="cagL"
          complement(5794..6507)
CDS      /gene="cagL"
          complement(5794..6507)

Query Match      100.0%; Score 685; DB 1; Length 20229;
Best Local Similarity 100.0%; Pred. No.3.3e-118;
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

gene
CDS
1 AAAAATGGCAAAATAGATTTACAGAGGTACCGCAAGCAAAAGCCGCTTGAAT 60
17085 AAAAATGGCAAAATAGATTTACAGAGGTACCGCAAGCAAAAGCCGCTTGAAT 17144
61 TCCGTAAAGATGTATCATCAATCAAAGGTACGATTAAGTTGATTAATCTCAATCAA 120
17145 TCCGTAAAGATGTATCATCAATCAAAGGTACGATTAAGTTGATTAATCTCAATCAA 17204
121 GCGGTATCAGTGGCTAAAGCAAGGGGTGATTTCAAGAGGTGAGAGCAAGCTTAAGCCGAT 180
17205 GCGGTATCAGTGGCTAAAGCAAGGGGTGATTTCAAGAGGTGAGAGCAAGCTTAAGCCGAT 17264
181 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAAGCTCAAAAAAATGAAGTCTCAAT 240
17265 CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAAGCTCAAAAAAATGAAGTCTCAAT 17324
241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGTGAATGGAACCTTAGTC 300
17325 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGTGAATGGAACCTTAGTC 17384
301 GGTAAATGGGTATCTCAAGAGCAAGCAAGCTTTCTTAAAAAATTTTCGACATCAAG 360
17385 GGTAAATGGGTATCTCAAGAGCAAGCAAGCTTTCTTAAAAAATTTTCGACATCAAG 17444
361 AAAGAGTTAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAACAATTAACAAT 420
17445 AAAGAGTTAATGCAAACTTGAAATTTCAATTAACAATTAACAATTAACAATTAACAAT 17504

```

```

Qy      421 GAACCAATTATGCTAAAGTTAATAAAGAAAGCAAGGCGACAGCTAGCTTGAAGA 480
Db      17505 GAACCAATTATGCTAAAGTTAATAAAGAAAGCAAGGCGACAGCTAGCTTGAAGA 17564
Qy      481 CCATTTAGCGCTCAAGTGTGCTAAAGTAATGCAAAAATGACCGACTCAATCAATA 540
Db      17565 CCATTTAGCGCTCAAGTGTGCTAAAGTAATGCAAAAATGACCGACTCAATCAATA 17624
Qy      541 GCAAGTGGTTGGGTGTGTAGGCGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 600
Db      17625 GCAAGTGGTTGGGTGTGTGTAGGCGCAAGCAGCGGCTTCCCTTTGAAAAGCATGATAA 17684
Qy      601 GTTAGATGCTCAGTAGTAGGCGCTTCAAGCAATCAAGAAATGGCTCAAGAAATTGAC 660
Db      17685 GTTAGATGCTCAGTAGTAGGCGCTTCAAGCAATCAAGAAATGGCTCAAGAAATTGAC 17744
Qy      661 AATCTCAATCAAGCGGTATCAGAG 685
Db      17745 AATCTCAATCAAGCGGTATCAGAG 17769

RESULT 9
AB015416      3444 bp      DNA      linear      BCT 13-JUN-2001
LOCUS      Helicobacter pylori caga gene, complete cds, strain NCTC11637.
ACCESSION      AB015416
VERSION      AB015416.1 GI:12225025
KEYWORDS      caga.
SOURCE      Helicobacter pylori;
ORGANISM      Helicobacter pylori
REFERENCE      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
AUTHORS      Helicobacteriaceae; Helicobacter.
1 (sites)
Hoshino, F.B., Katayama, K., Watanabe, K., Takahashi, S., Uchimura, H.
and Ando, T.
Heterogeneity found in the caga gene of Helicobacter pylori from
Japanese and non-Japanese isolates
J. Gastroenterol. 35 (12), 890-897 (2000)
11573724
2 (bases 1 to 3444)
Hoshino, F.B.
Direct Submision
Submitted (09-JUN-1998) Fuminoiri B Hoshino, Biomedical
Laboratories, (In: Basic Research Division, 1361-1 Maroba, Kawagoe,
Saitama 350-1101, Japan (E-mail: boone3@alk.co.jp,
Tel:81-492-32-0440, Fax:81-492-32-5480)
FEATURES
source
1..3444
location/Qualifiers
1..3444
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="NCTC 11637"
/db_xref="taxon:210"
1..3444
/gene="caga"
1..3444
/gene="caga"
1..3444
/codon_start=1
/transl_table=11
/product="Caga"
/protein_id="BAB20926.1"
/db_xref="GI:12225026"
/translation="MTNETIDQOPTEAFNPQPFINNLOVAFLKVDNAVASVDPDQK
SIVKNDNRNOAFEGISQLBEYSNKAIKPTKNOYFSDFINKSNLIDKNDLIDV
ESSTKSPKPDQVRIETSVSHQNDPSKNTNISRPMENITOPPLIDDKKAEFL
KSARSPGIIITGNQITDQKFMGVFDESLERQAEKNGSPGTGWDLIDFLSTFDK
KQSDVKEAINQEPVHPQDPIATTTTDDIQGLPEBARDLDERGAFKFTLDMEMLD
VEGYADIDPNYKFNQLIHNALSSVLAGSRNGIEPBEVSLIYGGNGGPGARHWNAT
VGYDQOQGNVATLITNVKNGSGLVLAGSRGGINNPSFYLYKEDQGLGSRALSOE
IONKIDFMEFLAONNAKLDNLSEKKEKFRTEIKDPODSKAYVLDALGNDAIATVSK
DTKSAITLTERGNDDISYTLKDYGRKADKLDREKNTVLOSLSLGDGMFVDYNSFKY
TNASKNPKGCVATNGVSHLEVGKRAIATFPLDNLNATLTSFVRNLEDDLTGKLS
POEANKLIDFLSSNKBELVGTILNPNNAVADAKNTGVDEVKQAKOLEKSLRRREHL

```

UB-09-360-685c-26.rpt

Page 8

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	YEAR	PAGES
Helicobacter pylori						
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;						
Helicobacteraceae; Helicobacter.						
(Pages 1 to 346)						
Knorr, R., Hartig, R., Konig, W., and Bacter, S.						
NE-Kamp, B.						
Identification and Potentiation of Proinflammatory						
Protein by the Helicobacter pylori CagA Protein						
Proc. Natl. Acad. Sci. U.S.A.						
15972330						
102 (26), 3500-3505						
(2005)						

FEATURES

University, Leipzig
Location/Qualifiers
Magdeburg 39120, Germany
Otto von Guericke University, Medical Microbiology,
Leipzig, Germany

CDS

```

/codon_start=1
/tranl_table=11
/product="CAGA"
/protein_id="AAV18597.1"
/db_xref="GI:62871423"
/tranl_table="http://www.ncbi.nlm.nih.gov/

```

PVLDKNDNRDRAPEKISOLBEEFPAKKAIPITKQVYDFDEIKSDNLDKNDKNDLADV
 BSGTSTKQKQKQORARITRITVSHQNDSEKNTKSTIRNEMNILOPIIPDDKKEFL
 KRAKOSRAGIITGNDIRTDQKFWGDESLKEBERKNGSGTGDMLDITPLSPFDK
 KOSSDVAGALINSEPBHQPDIATYTTDIOJLPEARADLDERKSGCFKFDKNDMD
 VAGADIDPEYKFNQDLHNNAALSYVLGMSNGIJEPEKSTIYGCNGCPBZHMND

QOUPMEFLAQVTLVRLKHKRSGUJYVGGEGGKNSRFTLYVEKQMGMTLALSOKEK
HIALVAFNGAGLSTYLKQKGRKODALBRETFTYQGLKHDGEMVAVNSRPYTKA
SKSDSDVAGVATGVSILAEANI,SKVAFNPLPIYHMLHITSYLRBDLQKIRAGLSFQJ
ANKLIXOPJNSTKELVEKALINPKFTVAAKQNSNDVDEKQDLEKSLJRGHEJLEK
VYKLEKRNKNGKREAKQANSOKDKFTATNPGAGTADN

ALBENIANSKDDKDSPEFENGKNDMPADRTTAKLREERFORSOMPEKENTLSEB
LNEFKGKDKDSKYTOAKDLENSIKDVIINOKTIDKDDYNDQANSIAKGSDSESV
BOALADIKNFSKQALAOAJRDNENFGSKSELXOSYVDQNTLTVGNLSISITATL
TKNSDLSKLGKLTKNFNNNNNNNLKNPEPIYAKVAKKTKTOSLEEBIYVOVAKY
NAKIDRNOITAGGKNGVSEPIJAKHDKDLSIKVGLSNQSLBLQKIDINDINQAVE
KAKPEPNTFACTGKNGVSEPIJAKHDKDLSIKVGLSNQSLBLQKIDINDINQAVE

Query Match	Score	DB 1	Length
87.9%;	601.8;	DB 1;	3426;
Best Local Siml:art...	0		
<p> KNGINEKKTIGTLQKRPPEMLTAVNDIKYKVAHNGAFPSLSKRLTAVINSHKLRINSKI KFTKRLNNAVVDIISGTFQFLINTSTASTICLAEENAKHGIKNANTKGFQKS" </p>			

Matches 633; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Y

1 AAAAAATGCGAAAATTAGCATTTACGCAAGTAACGCAGCAAAAGCAGCTGAATAAT 60
2233 AAAAATGGCAAAATTAAGATTTACCAAGTAGTAAATCAATCAATTCCTTTCTTTTT 600

61 TCCGTAAAGATGATCATCATCAAAAGTAACGGATTAAGTTGATTAATCTCAATCA 120
2293 TCCATTTAAAGATGATCATCATCAAAAGTAACGGATTAAGTTGATTAATCTCAATCA 2352
121 GCGGATCACTGCGTAAGGTAACGGATTAAGTTGATTAATCTCAATCA 2352

2353 GGGGATCATAGTAAAGGACACGCCATTTCAGTAGGCTGACCAAGCGGTACCAGAT 2412
181 CTCAAAATTTCTAAGAGCATTGCCTCCACAAGCTCAAAAAAATGAAGTCAAT 240
2413 CTATGACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT

241 GCTGAAAAATCTGAATATATCATCATCGTTAGAGATGCTGGAATGGAACCTGTGTC 300
|||||
2473 ACTGGAAAAATCTGAACTATACCAACCGTTAGAGATGCTGTAATATGAAACCTGTGTC 2532

..... 36

```

|||||
Db 2533 GGTAAATGGGTTATCTGGAAATAGAGCCACAGCTCTCAACAAAATTTTTCGATATCAAG 2592
Qy 361 AAGAGTGAATCCAAAACCTTGCAAAATTTCAATTAACAATAACAATAATGACATCAAAAC 420
Db 2593 AAGAAATTTGATGAGAACTTAAATAATTTCAATTAACAATAACAATAATGAGCTCAAAAC 2652
Qy 421 GAACCAATTTATCTAAAGTTAATTAATAAAGAAAGAGGAGGAGAGCTAGCTTGAAGA 480
Db 2653 GAACCAATTTATCTAAAGTTAATTAATAAAGAAAGAGGAGGAGCTAGCTTGAAGA 2712
Qy 481 CCCATTAGAGCTCAAGTTGCTTAAAGAAAGTAATGCAAAATTTGACGATCAATCAAA 540
Db 2713 CCCATTATATCTCAAGTTGCTTAAAGAAAGTAATGCAAAATTTGACGATCAATCAAA 2772
Qy 541 GCAAGTGGTTGGGCTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCTGATAA 600
Db 2773 GCAAGTGGTTGGGCTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGCTGATAA 2832
Qy 601 GTTATGATCTCAGTAAGTAGAGGCTTTCAAGGATCAAGATTTGGCTCAGAAAATTGAC 660
Db 2833 GTTATGATCTCAGTAAGTAGAGGCTTTCAAGGATCAAGATTTGGCTCAGAAAATTGAC 2892
Qy 661 AATCTCAATCAAGCGGTATCAGAG 685
Db 2893 AATCTCAATCAAGCGGTATCAGAG 2917

RESULT 11
AB110963 978 bp DNA linear BCT 23-SEP-2004
LOCUS Helicobacter pylori caga gene for Caga, partial cds, strain:
DEFINITION China:079a.
ACCESSION AB110963
VERSION AB110963.1 GI:31339271
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteriaceae; Helicobacter.
REFERENCE
AUTHORS Zhou, W., Yamazaki, S., Yamakawa, A., Ohtani, M., Ito, Y., Keida, Y.,
Higashi, H., Hatakeyama, M., Si, D., and Azuma, T.
TITLE The diversity of vacA and caga genes of Helicobacter pylori in East
JOURNAL Asia
PUBMED FEMS Immunol. Med. Microbiol. 40 (1), 81-87 (2004)
AUTHORS 14734191
2 (bases 1 to 978)
Zhou, W., Yamazaki, S., Yamakawa, A. and Azuma, T.
Direct Submission
Submitted (29-MAY-2003) Takeshi Azuma, Fukui Medical University,
Second Department of Internal Medicine, Matsuoka-cho, Yoshida-gun,
Fukui 910-1193, Japan (E-mail: azuma@fmsra.fukui-med.ac.jp,
Tel: 81-776-61-8351, Fax: 81-776-61-8110)
FEATURES
source 1. 978
Location/Qualifiers
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="China:079a"
/db_xref="taxon:210"
/country="China"
/note="Isolated from a patient with duodenal ulcer"
1..978
/gene="caga"
<1..>978
/gene="caga"
/codon_start=1
/transl_table=11
/product="Caga"
/protein_id="BAC71006.1"
/db_xref="GI:31339272"
/translation="LKDFKSFDEFKNGKNDKPSKTEETLKALXDSVQDLGNPEWIS
KVENITALEFKNGKNDKPSKVTQAKSLKENSIDVLIINQKITDKNDINQAVSIAR

```

ANGPSGVEOVLAIDIKNPSKGLAQAOKNEDPNTKNSLQSYKNGVNGTLVNGEL
 SQAEATLTSKNSIDIKKEINERKXFNPNNNNGIKNEPIYAAVNNKRGVNSLEPI
 YTOVANNVAKIDRLNQLASGFGVGAQFPLKDKHDKVGLDLSKGLSBEQSLKOKID
 NINQAVSEAKGHFDNLVOMIDELDKSTKQNVNIMVESAKKVPASLSAK"

Query Match 87.2%; Score 597; DB 1; Length 978;
 Best Local Similarity 92.0%; Pred. No. 2,6e-101;
 Matches 630; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```

Qy 1 AAAAAATGGCAAAAATTAAGATTTTCAGCAAGTAAGCAAGCAAAAAGCGACCTGTAAT 60
Db 169 AAAAAATGGCAAAAATTAAGATTTTCAGCAAGTAAGCAAGCAAAAAGCGACCTGTAAT 228
Qy 61 TCGTTAAAGATGTGATCATCATCAAAAAGTAACGGATTAAGTTGATATCTCAATCAA 120
Db 229 TCGATTAAAGATGTGATCATCATCAAAAAGTAACGGATTAAGTTGATATCTCAATCAA 288
Qy 121 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTTTCAGTGGGTAGACCAAGCTTACGCCAT 180
Db 289 GCGGTATCAGTGGCTTAAAGCAACGGGTGATTTTCAGTGGGTAGACCAAGCTTACGCCAT 348
Qy 181 CTCAAAAATTTCTCAAAAGAGCAATTTGGCCCAAGAGCTCAAAAAATGAAGTCTCAAT 240
Db 349 CTCAGAAATTTCTCAAAAGAGCAATTTGGCTCAACAGCTCAAAAAATGAAGTCTCAAT 408
Qy 241 GGTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGTGTGATGAAACCCCTAGTC 300
Db 409 ACTGAAAAAATTTCTGAACATATACCATCCGTTAAGAAATGTGTGATGAAACCCCTAGTC 468
Qy 301 GGTATAGGGTTATCTCAGACAGACCAACTCTTTCTAAACCTTTGGACATCAAG 360
Db 469 GGTAAAGGGTTATCTCAGACAGACCAACTCTTTCTAAACCTTTGGATATCAAG 528
Qy 361 AAGAGTTGAATGCAAAAATTTGAAATTTCAATAACAATAACAATAAGTAAGTAAGTAAG 420
Db 529 AAGAGTTGAATGCAAAAATTTGAAATTTCAATAACAATAACAATAAGTAAGTAAGTAAG 588
Qy 421 GAACCAATTTATGCTAAAGTTAATTAAGAAAGAGGAGGAGCAAGCTAGCTTGAAGA 480
Db 589 GAACCAATTTATGCTAAAGTTAATTAAGAAAGAGGAGGAGCAAGCTAGCTTGAAGA 648
Qy 481 CCCATTAGAGCTCAGTAGGCTTAAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 540
Db 649 CCCATTATATCTCAAGTTGCTTAAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 708
Qy 541 GCAAGTGGTTGGGCTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGATGATAA 600
Db 709 GCAAGTGGTTGGGCTGTGTGAGGCAAGCAGCGGCTTCCCTTTGAAAAGGATGATAA 768
Qy 601 GTTATGATCTCAGTAAGTAGAGGCTTTCAAGGATCAAGATTTGGCTCAGAAAATTGAC 660
Db 769 GTTATGATCTCAGTAAGTAGAGGCTTTCAAGGATCAAGATTTGGCTCAGAAAATTGAC 828
Qy 661 AATCTCAATCAAGCGGTATCAGAG 685
Db 829 AATCTCAATCAAGCGGTATCAGAG 853

RESULT 12
LOCUS D0067454 3693 bp DNA linear BCT 15-JUN-2005
DEFINITION Helicobacter pylori strain 7bgs cytotoxin-associated protein A
ACCESSION D0067454
VERSION D0067454.1 GI:67483340
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteriaceae; Helicobacter.
REFERENCE
AUTHORS Kim, S.Y., Blaser, M.J., Lee, Y.C. and Pillinger, M.H.

```


TITLE Helicobacter pylori stimulates matrix metalloproteinase-1 secretion from gastric epithelial cells via CagA-dependent and independent mechanisms: requirement for Etx activation

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3693)

AUTHORS Kim, S. Y., Blaser, M. J., Lee, Y. C. and Pillinger, M. H.

TITLE Direct Submission

JOURNAL Submitted (18-MAY-2005) Microbiology, College of Medicine, Chungbuk National University, San 48, Gaesin-dong, Heungduk-ku, Cheongju, Chungbuk 361-763, Korea

FEATURES

source

1..3693

/organism="Helicobacter pylori"

/mol_type="genomic DNA"

/strain="7Dqs"

/db_xref="Caxon:210"

1..3693

/gene="CagA"

1..3693

/gene="CagA"

/note="CagA"

/codon_start=1

/transl_table=11

/product="cytotoxin-associated protein A"

/protein_id="AA68033.1"

/db_xref="GI:67483341"

/translation="MTNETINQOPQTEAFNPQFNNLRAPFKYNNVAAPSDPDKPIVDKNDNRNDAFDGI SOLREYSNKAIKNPKKNQYFSSPIKSNLDLNNCLIDYDSYSPQKFGDQRYOIFMDWVSHQNDPSKINTRSIRNPMNITOPRIDPDKLELRSKQSFAGIIGNQIRTDQKFMGVDPESIKERQDAERTGDMIDITSPIDTKQSSDIKETLNOEPVHVDIATTTDIOGLPEARADLLDRGNSPSTGICPMKEDLVGADIDENYKFNQOLIHNNAISVLSGNSNGVPEKXSLILYGGNGSPRABRNDALVIGVDOGNNAITINVMKNGSGVIVAGSEKGIINPSPYLKKNQGLNRSQSEIRNKVDMEFLKNNALDNLSEKEREKETEIDQKDSKAYLDLGNRLNRSKQPNKSAIITERNKDLSTYTKDYKKKADKALDRKRVITRGILKRDQMVVDLSNRTKTNKSNPKNGVATNGVSHLDAGFSKAVFPLPILNNLAIITSVARDELKRLNAGLSLOEVNKLIDPLNSNKELVGLYLNKAVANKEVNTGNDVRYKQKQDLSEKLSKEHLEKVEKLESKSNKMKMEAKIQANSODEIPALIKKEANDRAIATQNLKIGIKRELSDKLENNKRLKDPKSPDEPNKNDPSKENTKIKKSKSVADGIVNPIKSVENLDAVEOLADLKVFSKEQLAQAKNESPNVKEEITQSVKNGVGLVNGSKAEATFLSKSPDIKKELNAGLGNFNNNNNGIKNEPIITAVNKKAGALSEPIYAAVAAKVNKIDRLNQLASGLGVQAGPPIKRDNRVDKQDSKSVSPSEPIYATIDLGPPPLKRDHDKVQLSKVGSRNQKLAOKTIDNDAVSEKSGFSPEPIYATIDLGPPPLKRWESAKKVPASISATLDNVATNSHTKLNININHGAIKEATYMTQKPEMLKLVNDKIVAHNSGAPLISAYDKIGPDNRNMDYSDSPFSETKLNNAVNDIKRFTOFLANACSTGYCLAEENAHGIDINVTIKGFQKS"

ORIGIN

Query Match 86.9%; Score 595.6; DB 1; Length 3693;

Best Local Similarity 95.5%; Pred. No. 2.8e-101;

Matches 613; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

1 AAAAAATGCAAAAATTAAGATTTTCAGCAAGTAAACGCAAGCAAAAGCGACCTTGAAAT 60

2236 AAAAAATGCAAAAATTAAGATTTTCAGCAAGTAAACGCAAGCAAAAGCGACCTTGAAAT 2295

61 TCCGTTAAAGATGTCATCATCAAAAAGTAAAGGATTAAGTGAATATCTCAATCA 120

2296 TCCATTAAGATGTCATCATCAAAAAGTAAAGGATTAAGTGAATATCTCAATCA 2355

121 GCGGTATCATGTCCTTAAGCAAGGATTTTCAGTAAAGTAAAGCAAGCGTTAGCCGANT 2415

2356 GCGGTATCATGTCCTTAAGCAAGGATTTTCAGTAAAGTAAAGCAAGCGTTAGCCGANT 2415

181 CTCGAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTCAAAAAAAGAAAGTTCAT 240

2416 CTCGAAAAATTTCTCAAGAGCAATTTGCGCAACAGCTCAAAAAAAGAAAGTTCAT 2475

241 GCTGAAAAAATCTGAATAATATCATCCGTTAAAGATGTTGAATGTAAGCAACCTTATG 300

2476 GCTGAAAAAATCTGAATAATATCATCCGTTAAAGATGTTGAATGTAAGCAACCTTATG 2535

QY 301 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAAATTTTGGACATCAAG 360

Db 2536 GGTAAATGGTTATCTCAAGCAAGCAAGCAACTCTTTTAAAAAATTTTGGACATCAAG 2595

QY 361 AAGAGTGAATGCAAAACCTTGAAATTTTCAATTAACATTAATGATGACTCAAAAC 420

Db 2596 AAGAGTGAATGCAAAACCTTGAAATTTTCAATTAACATTAATGATGACTCAAAAC 2655

QY 421 GAACCATTTATGCTAAAGTTAAATTAATAAAGAGGAGGAGGAGGAGGAGGAGGAGG 480

Db 2656 GAACCATTTATGCTAAAGTTAAATTAATAAAGAGGAGGAGGAGGAGGAGGAGGAGG 2715

QY 481 CCCATTTACGCTCAAGTTGCTAAAGTTAAATTAATAAAGAGGAGGAGGAGGAGGAGG 540

Db 2716 CCCATTTACGCTCAAGTTGCTAAAGTTAAATTAATAAAGAGGAGGAGGAGGAGGAGG 2775

QY 541 GCAAGTGTGTTGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600

Db 2776 GCAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2835

QY 601 GTTATGATCTCAGTAAAGTAAAGGCTTTCAAGCAATCAAGA 642

Db 2836 GTTATGATCTCAGTAAAGTAAAGGCTTTCAAGCAATCAAGA 2877

RESULT 13

AY3063882 26663 bp DNA linear BCT 25-MAR-2004

LOCUS Helicobacter pylori strain Car3 cag pathogenicity island section 2, AY306382

DEFINITION Partial sequence.

ACCESSION AY30639

VERSION AY30639.1 GI:37811810

KEYWORDS

SEGMENT

SOURCE

ORGANISM

2 of 2

Helicobacter pylori

Helicobacter pylori

Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.

1 (bases 1 to 26663)

Lundberg, A., Lundin, A., Nilsson, C., Engstrand, L. and Lundberg, J.

2 (bases 1 to 26663)

Comparative analysis of the complete cag pathogenicity island sequences in four Helicobacter pylori isolates

1501998, 85-93 (2004)

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

1501998

2 (bases 1 to 26663)

Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. and Lundberg, J.

Direct Submission

Submitted (26-JUN-2003) Biotechnology, Royal Institute of Technology, Alpha Nova University Centre, Roslagstullsbacken 21, S-106 91 Stockholm, Sweden

FEATURES

source

1..26663

/organism="Helicobacter pylori"

/mol_type="genomic DNA"

/strain="Ca73"

/db_xref="Caxon:210"

<1..25906

/note="cag pathogenicity island"

complement(<1..2031)

/gene="cag7"

/locus_tag="HP0476"

/note="cag-Y; jhp0476; orf13/14; vitB10-like protein"

/codon_start=1

/transl_table=11

/product="Cag7"

/protein_id="AA03890.1"

/db_xref="GI:37811818"

/translation="MTNETINQOPQTEAFNPQFNNLRAPFKYNNVAAPSDPDKPIVDKNDNRNDAFDGI SOLREYSNKAIKNPKKNQYFSSPIKSNLDLNNCLIDYDSYSPQKFGDQRYOIFMDWVSHQNDPSKINTRSIRNPMNITOPRIDPDKLELRSKQSFAGIIGNQIRTDQKFMGVDPESIKERQDAERTGDMIDITSPIDTKQSSDIKETLNOEPVHVDIATTTDIOGLPEARADLLDRGNSPSTGICPMKEDLVGADIDENYKFNQOLIHNNAISVLSGNSNGVPEKXSLILYGGNGSPRABRNDALVIGVDOGNNAITINVMKNGSGVIVAGSEKGIINPSPYLKKNQGLNRSQSEIRNKVDMEFLKNNALDNLSEKEREKETEIDQKDSKAYLDLGNRLNRSKQPNKSAIITERNKDLSTYTKDYKKKADKALDRKRVITRGILKRDQMVVDLSNRTKTNKSNPKNGVATNGVSHLDAGFSKAVFPLPILNNLAIITSVARDELKRLNAGLSLOEVNKLIDPLNSNKELVGLYLNKAVANKEVNTGNDVRYKQKQDLSEKLSKEHLEKVEKLESKSNKMKMEAKIQANSODEIPALIKKEANDRAIATQNLKIGIKRELSDKLENNKRLKDPKSPDEPNKNDPSKENTKIKKSKSVADGIVNPIKSVENLDAVEOLADLKVFSKEQLAQAKNESPNVKEEITQSVKNGVGLVNGSKAEATFLSKSPDIKKELNAGLGNFNNNNNGIKNEPIITAVNKKAGALSEPIYAAVAAKVNKIDRLNQLASGLGVQAGPPIKRDNRVDKQDSKSVSPSEPIYATIDLGPPPLKRDHDKVQLSKVGSRNQKLAOKTIDNDAVSEKSGFSPEPIYATIDLGPPPLKRWESAKKVPASISATLDNVATNSHTKLNININHGAIKEATYMTQKPEMLKLVNDKIVAHNSGAPLISAYDKIGPDNRNMDYSDSPFSETKLNNAVNDIKRFTOFLANACSTGYCLAEENAHGIDINVTIKGFQKS"

CDs

misc_feature

complement(<1..2031)

/gene="cag7"

/locus_tag="HP0476"

/note="cag-Y; jhp0476; orf13/14; vitB10-like protein"

/codon_start=1

/transl_table=11

/product="Cag7"

/protein_id="AA03890.1"

/db_xref="GI:37811818"

/translation="MTNETINQOPQTEAFNPQFNNLRAPFKYNNVAAPSDPDKPIVDKNDNRNDAFDGI SOLREYSNKAIKNPKKNQYFSSPIKSNLDLNNCLIDYDSYSPQKFGDQRYOIFMDWVSHQNDPSKINTRSIRNPMNITOPRIDPDKLELRSKQSFAGIIGNQIRTDQKFMGVDPESIKERQDAERTGDMIDITSPIDTKQSSDIKETLNOEPVHVDIATTTDIOGLPEARADLLDRGNSPSTGICPMKEDLVGADIDENYKFNQOLIHNNAISVLSGNSNGVPEKXSLILYGGNGSPRABRNDALVIGVDOGNNAITINVMKNGSGVIVAGSEKGIINPSPYLKKNQGLNRSQSEIRNKVDMEFLKNNALDNLSEKEREKETEIDQKDSKAYLDLGNRLNRSKQPNKSAIITERNKDLSTYTKDYKKKADKALDRKRVITRGILKRDQMVVDLSNRTKTNKSNPKNGVATNGVSHLDAGFSKAVFPLPILNNLAIITSVARDELKRLNAGLSLOEVNKLIDPLNSNKELVGLYLNKAVANKEVNTGNDVRYKQKQDLSEKLSKEHLEKVEKLESKSNKMKMEAKIQANSODEIPALIKKEANDRAIATQNLKIGIKRELSDKLENNKRLKDPKSPDEPNKNDPSKENTKIKKSKSVADGIVNPIKSVENLDAVEOLADLKVFSKEQLAQAKNESPNVKEEITQSVKNGVGLVNGSKAEATFLSKSPDIKKELNAGLGNFNNNNNGIKNEPIITAVNKKAGALSEPIYAAVAAKVNKIDRLNQLASGLGVQAGPPIKRDNRVDKQDSKSVSPSEPIYATIDLGPPPLKRDHDKVQLSKVGSRNQKLAOKTIDNDAVSEKSGFSPEPIYATIDLGPPPLKRWESAKKVPASISATLDNVATNSHTKLNININHGAIKEATYMTQKPEMLKLVNDKIVAHNSGAPLISAYDKIGPDNRNMDYSDSPFSETKLNNAVNDIKRFTOFLANACSTGYCLAEENAHGIDINVTIKGFQKS"

/db_xref="GI:37811826"
/translation="MTLVYPLFLPLFLIKSPFYTYTLLLVLSLISIKYLLSHA
NFTKLLIATOMOMFL"
gene
10724..11854
/locus_tag="HPO537"
10724..11854
/gene="cag16"
/locus_tag="HPO537"

Query Match
Best Local Similarity 94.9%; Score 589.2; DB 1; Length 26663;
Matches 609; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 AAAAATGCAAAAATAGATTTCAGACGAGTAAAGCAAGCAAAAAGCAAGCTTGAAT 60
DB 24161 AAAAATGCAAAAATAGATTTCAGACGAGTAAAGCAAGCAAAAAGCAAGCTTGAAT 24220
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGTAAAGTAAATCATCAATCA 120
DB 24221 TCCATTAAGATGTATCATCAATCAAAAGTAAAGTAAAGTAAATCATCAATCA 24280
QY 121 GCGGATATCATGTGCTAAAGCAAGCGGTGATTCAGTAAAGCAAGCAAGCGGTGAT 180
DB 24281 GCTGTATCATGTGCTAAAGCAAGCGGTGATTCAGTAAAGCAAGCAAGCGGTGAT 24340
QY 181 CTCAAAATTTCTCAAAAGCAATTTGGCCCAAGCTCAAAAATGAAGTCAAT 240
DB 24341 CTCAAAATTTCTCAAAAGCAATTTGGCCCAAGCTCAAAAATGAAGTCAAT 24400
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGATTAAGCACTTGAAT 300
DB 24401 ACTGAAAAAATCTGAATATATCAATCCGTTAAGATGTGATTAAGCACTTGAAT 24460
QY 301 GGTATAGGTGTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360
DB 24461 GGTATAGGTGTATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 24520
QY 361 AAAGAGTGAATGCAAAAATCTGAATTTCAATCAATCAATCAATCAATCAATCA 420
DB 24521 AAAGAGTGAATGCAAAAATCTGAATTTCAATCAATCAATCAATCAATCAATCA 24580
QY 421 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 480
DB 24581 GAACCAATTTATGCTAAAGTAAATTAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 24640
QY 481 CCCATTAGCGTCAAGTGTGCTAAAGTAAATGAAGTAAATGAAGTAAATGAAGTAA 540
DB 24641 CCCATTAGCGTCAAGTGTGCTAAAGTAAATGAAGTAAATGAAGTAAATGAAGTAA 24700
QY 541 GCAAGTGTGTTGGGTGTGTTAGGCGCAAGCAAGCGGTCTCCCTTGAAGCAAGTAA 600
DB 24701 GCAAGTGTGTTGGGTGTGTTAGGCGCAAGCAAGCGGTCTCCCTTGAAGCAAGTAA 24760
QY 601 GTTGAATATCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 642
DB 24761 GTTGAATATCTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 24802

RESULT 14
AB090088 3543 bp DNA linear BCT 07-NOV-2002
DEFINITION Helicobacter pylori gene for CagA, complete cds, strain:OK12.
ACCESSION AB090088
VERSION AB090088.1 GI:22335784
KEYWORDS
SOURCE Helicobacter pylori
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 Higashi, H., Tsubota, R., Fujita, A., Yamazaki, S., Aaka, M., Azuma, T.
and Hatakeyama, M.

TITLE
Biological activity of the Helicobacter pylori virulence factor
CagA is determined by variation in the tyrosine phosphorylation
sites
Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14428-14433 (2002)
JOURNAL
PUBMED
12391297
AUTHORS
Yamazaki, A., Yamazaki, S. and Azuma, T.
TITLE
JOURNAL
Submitted (19-AUG-2002) Takeshi Azuma, Fukui Medical University,
Second Department of Internal Medicine, Matsuoka-cho, Yoshida-gun,
Fukui 910-1193, Japan (E-mail: azuma@fmsr.fukui-med.ac.jp,
Tel:81-776-61-8351, Fax:81-776-61-8110)
FEATURES
source
1..3543
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="OK12"
/db_xref="taxon:210"
/country="Japan"
1..3543
/gene="caga"
1..3543
/gene="caga"
/product="CagA"
/protein_id="BAC10434.1"
/db_xref="GI:22335785"

ORIGIN
Query Match
Best Local Similarity 94.4%; Score 584.4; DB 1; Length 3543;
Matches 606; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 1 AAAAATGCAAAAATAGATTTCAGACGAGTAAAGCAAGCAAAAAGCAAGCTTGAAT 60
DB 2248 AAAAATGCAAAAATAGATTTCAGACGAGTAAAGCAAGCAAAAAGCAAGCTTGAAT 2307
QY 61 TCCGTTAAGATGTATCATCAATCAAAAGTAAAGTAAAGTAAATCATCAATCA 120
DB 2308 TCCATTAAGATGTATCATCAATCAAAAGTAAAGTAAAGTAAATCATCAATCA 2367
QY 121 GCGGATATCATGTGCTAAAGCAAGCGGTGATTCAGTAAAGCAAGCAAGCGGTGAT 180
DB 2368 GCGGATATCATGTGCTAAAGCAAGCGGTGATTCAGTAAAGCAAGCAAGCGGTGAT 2427
QY 181 CTCAAAATTTCTCAAAAGCAATTTGGCCCAAGCTCAAAAATGAAGTCAAT 240
DB 2428 CTCAAAATTTCTCAAAAGCAATTTGGCCCAAGCTCAAAAATGAAGTCAAT 2487
QY 241 GCTAGAAAAAATCTGAATATATCAATCCGTTAAGATGTGATTAAGCACTTGAAT 300

Db 2488 ACTGAAAAAATTCGTGACATATACCAATCCGTTAAGATGTGTAAAAAGAACCCCTAGTC 2547
Qy 301 GGTAAATGGTTATCTGAAGAGAACCAACTCTTTCTAAAACTTTTGGGATCAAG 360
Db 2548 GGTAAATGGTTATCTGAAGAGAACCAACTCTTTCTAAAACTTTTGGGATCAAG 2607
Qy 361 AAAGAGTGAATGCAAAACTTGGAAATTTCAATACATATACATATATGACTCAAAAC 420
Db 2608 AAAGAGTGAATGCAAAACTTGGAAATTTCAATACATATACATATATGACTCAAAAC 2667
Qy 421 GAACCCATTATGCTAAAGTTATATTAAGAAAGAGGAGGACAGCAGCTTGAAGAA 480
Db 2668 GAACCCATTATGCTAAAGTTATATTAAGAAAGAGGAGGACAGCAGCTTGAAGAA 2727
Qy 481 CCCATTAGCGCTCAAGTGTGTTAAAGGTAATGCAAAATTTAGACGATCAATCAAA 540
Db 2728 CCCATTAGCGCTCAAGTGTGTTAAAGGTAATGCAAAATTTAGACGATCAATCAAA 2787
Qy 541 GCAAGTGTGTTGGTGTGTTAGGAGCAAGCAGCGGCTTCCCTTGAAGAAAGCATGATA 600
Db 2788 GCAAGTGTGTTGGTGTGTTAGGAGCAAGCAGCGGCTTCCCTTGAAGAAAGCATGATA 2847
Qy 601 GTTGATGATCTCAGTAAGGTAAGGCGCTTCAAGGATCAAGAA 642
Db 2848 GTTGATGATCTCAGTAAGGTAAGGCGATCGTTAGCCCTGAA 2889

RESULT 15
AY330636S2 26771 bp DNA linear BCT 25-MAR-2004
LOCUS Helicobacter pylori strain Cas2 cag pathogenicity island section 2,
DEFINITION partial sequence.
ACCESSION AY330637
VERSION AY330637.1 GI:37811777
KEYWORDS
SEGMENT
SOURCE
ORGANISM
2 of 2
Helicobacter pylori
Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
1 (bases 1 to 26771)
Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. and
Lundberg, J.
Comparative analysis of the complete cag pathogenicity island
sequence in four Helicobacter pylori isolates
Gene 328, 85-93 (2004)
15019987
2 (bases 1 to 26771)
Blomstergren, A., Lundin, A., Nilsson, C., Engstrand, L. and
Lundberg, J.
Direct Submision
Submitted (26-JUN-2003) Biotechnology, Royal Institute of
Technology, Alba Nova University Centre, Roslagstullsbacken 21,
S-106 91 Stockholm, Sweden
Location/Qualifiers
1.26771
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="Cas2"
/db_xref="taxon:210"
<1..26014
/note="cag pathogenicity island"
complement(1..2025)
/gene="cag"
/locus_tag="HP0527"
/note="cag-Y; jhp0476; orf13/14; vlrB10-like protein"
/codon_start=1
/product="Cag7"
/protein_id="AA03862.1"
/db_xref="GI:37811787"
/translation="MNEEDKLETSKKAQOHSPODLSNEETEAHPEDLLKESSDH
YLDNPLETQTNPDGDKSEETQYQMDSGNETSESSNSGLADLCLKKARKLVNKRPPPT

OOKSLDEFTOEINEEDDNNNEOEETQDLDIDETSCKAQSPODLSNEETEAHP
PDSLSKSSDNHLDNPLETQTNPDGDKSEETQYQMDSGNETSESSNSGLADLCLK
AKLVNKRPPPTQOXSLEDETOELNEEDQENNEYOETQDLDIDETSCKAQSPO
DLSNEETEAHPEDLSKSSDNHLDNPLETQTNPDGDKSEETQYQMDSGNETSESS
KXYIIGIVATLIVILFERSIFHFMPLDEKSSRSPKORALYNNDEIOIQOENRL
LKERNEKMMIDKNLEFNDPRLTYNNILAEIBDKNPLRAFYECISNGAYEBCIL
LKDKKLDOMKCTEAYVYDCKTKNAKTEBERIKCDLKDENLKSLNQOQYVALD
CLNAKTDBERRECKLINDPEIRKREKLEQOELDEYKOCITNAKTEBERCKL
GUSKALIAELKQALDCLKNKATDERNECKNITPODLQKELLAQMSVAYVDQVSKA
RNEKRECECKLITPEAKKLEQOVLDCIKNAKTDBERCKLIDPKLOS"
/gene="cag8"
/locus_tag="HP0528"
complement(2040..3608)
/gene="cag8"
/locus_tag="HP0528"
/note="cag-X; jhp0477; orf15; vlrB9-like protein"
/codon_start=1
/transl_table=1
/product="Cag8"
/protein_id="AA03863.1"
/db_xref="GI:37811788"
/translation="MEQAFPKIYVCGFCGLGYLFLSSVIEAAPDIKNFGRKVVNK
KIAYGDEKPIITWISLDNVTYIOLEKDETSYLTGPNKMSIYPNSNHIPIQKSV
KSNLAFKRAVNFALMTDYQFLTKCLIDAPPKLEBQKALBEKKEAKBOAQ
AOKDREKREERANRANLENTNAMSNPOLSNKKLSEFTKOORENELOMERLE
DMQEOQANALKOIELNKQAEFTIKORAKDKSITKDKQSPEDSISLPSDSA
WRTNLVVRNTKALYQFLIRIAQDNFASAYLVKLEYPQREHVSIVIEELKREBA
RORELIKOENLNTATYINRWMASSNEOINNEKIREEKOKIILDOAKLETOYVNAL
KNPVPRAVNYVOAPEKRSKHIMSEIIPDDCTFYRGKNTITQPAIVVQPDGLSM
TDAIDPMKTSGLMWRVNEIAERFKLIKQALVTVINKGKNPILKTNVKNYGE
LERVILKULPVDRK"
complement(3661..5268)
/gene="cag9"
/locus_tag="HP0529"
complement(3661..5268)
/gene="cag9"
/locus_tag="HP0529"
/note="cag-W; jhp0478; orf16; vlrB8-like protein"
/codon_start=1
/transl_table=1
/product="Cag9"
/protein_id="AA03864.1"
/db_xref="GI:37811789"
/translation="MFNIKRTFLITIIISFPIVPMWLKAIDLPIVSNKIYQVYCM
IPSYLVNKSFPADILITGYISGASGSGSGQVLEALSTPLATSLASNNVKNYNTL
GPLMSANASAVATMIOGFALTPSSGSGNGMNAIKNNDVMSVDLNTSNKIONFTK
GSVGVNKNYNTLLOISGITQASNTITADGLIMLIGKFTANRKLQNTTANLAPAL
ESYVKGADAVLPAIGVNVLPDIIIGQSYLDPVSILYIVGIFVPLISPMKLRDSS
NIQINIGEYMRPVGGLFKMAWSFLAVAGFYLYKISYISYIFGLAGAFGLNOLFW
ALDVLANTVNSILPAPRAVSNVGNNAVSLLOGIQVAGISLFAIPMQVTILMRISTV
VVKPLIAGFSGIYVPIAVCLVIDMFDSKMNLIIMFINNLFTILVILPILFGLVLA
LLAPVLITTPSAVITONOGGIGDSTIASLITLPIAKGPIETIIESVNAIVNTIFSS
VSMGSRDRDRDALMVRGSGSMFKG"
complement(5273..6031)
/gene="cag10"
/locus_tag="HP0530"
complement(5273..6031)
/gene="cag10"
/locus_tag="HP0530"
/note="cag-V; jhp0479; orf17"
/codon_start=1
/transl_table=1
/product="Cag10"
/protein_id="AA03865.1"
/db_xref="GI:37811790"
/translation="WLGKKNEEVLIDENLVGVIATLDRLANLRANRTKRAFYLMA
LNVAVTSIVMMPLKKTIDIPVGIIDRTGFKLVKSDARQIVNSEAVVDSATSKFV
SILFGYSKNSLRDKDQMOYCDVSPQOARMENENIRQVDKVRABALISSNIORE
KVNSPLRLPFLPITIKTPTDMENVEITKKQVYIYDFARGNSQENMLINPFGFK
VFDIOTIDONEQVSHIRIKIKVESKNALNK"
6409..7059
/gene="cag11"

CDS

```

CDS
        /locus_tag="HP0531"
        6409..7059
        /gene="caq11"
        /locus_tag="HP0531"
        /note="caq-U; jhp0480, orf18"
        /codon_start=-1
        /transl_table=-1
        /product="Cag11"
        /protein_id="AA03866.1"
        /db_xref="GI:37811791"
        /transl_table="MDITTEHHGPNLPNPSNSQNDLLNTLSDSYERKSLGQFPF
        HEALANDEEKLILLEIKRLPOEKRYELLMOQGOAKORIMAAVAVGAVLFTGLSEHLL
        GLCWGSLFVLEALYFPFLRTKNDLVESLTKRLGLQFYAAVGAVAVLFTGLSEHLL
        IYELIYLGISLNDKLVLEVLNLCIFGMDGLYIKVITIKVIGKNPGLTSKQEQ"
        7072..71914

```

CDS

```

/gene="Cag12"
/locus_tag="HP0532"
/notes="cagT, jhp0481, of19; virB7-like protein"
/codon_start=1
/translate_table=11
/product="Cag12"
/protein_id="g12"
/db_xref="g:37811792"
/db_xref="g:37811792"
/translation="MKLASAVLTGATITCLILTSACSNVAKVKVKQNHVYTPVYNELI
EKSEPLPDKDKITPRMVQVCLKPNKYULLDNKQVLLPKVLHNSKSLTLIGPANKI
LOKRTFQNGKSGSIDFYLOPTLNQKGVMIASNYINDNNKNEKPTFVLSQSPM
LGATKLNHLDGVSANNQKQVINVARAEQLEKINQYITLLDQKQETYNRNR
ELIETLSNAGYQNRQVIVSSEIFKNGNLNQAAEEBVEVKIQEERENEVLRNQINSL
LSGR"
complement(8116..8715)

```

CDS

```

/gene="cag13"
/locus_tag="HP0534"
/notes="cag8; jhp0482; orf20"
/codon_start=1
/translation=1
/product="cag13"
/protein_id="AAK03668.1"
/db_xref="GI:37811793"
/translation="MSNNRKLFSMIADSKDKTEKLIIESIENELNTDEKKIITDOI
KTHDFPKQMHNTKAGLDVLYLNMYKQYRAVIGISIVDFKFKVRLLESEFTWELHAI
TENPFLPQKHTNIGLIFLPFSPKPIWFGNSLRENDSGIELVETGTLPLAKLPFWTDE
ENKFPANLKITEQYNDVYDLMLDEKTEKTKMEKGQVNA"
complement(9153..9533)

```

IDS

```

/gene="Cag14"
/locus_tag="HP0535"
/note="CagQ: jhp0463"
/codon_start=1
/translation_table=11
/product="Cag14"
/protein_id="AA03869.1"
/db_xref="GI:37811794"
/feature_location="MLPTKTRIDRNPKQELTOPKIKGLIMKIKLILASLGGGNTLPTQLSSDLFSFSLINLFFPLMLMGLENALCKEKNLPMNIIKAFNAEVLVKNQFDSIKNNGSMLVGNFSGSSSKDKTKANMVSQVRL"
/complement(9966..10310)
/gene="Cag15"
/locus_tag="HP0536"
/complement(9966..10310)

```

```

/codon_start=1
/transl_table=11
/product="Cag15"

```

```

/protein_id="AA03870.1"
/db_xref="GI:37811795"
/translation="MKRISLKNFLOPKHSFNHGDKSLVLPENISGVGFI
ALFSGVGVVLVPIFLFALLIYPSFFYTYLLLVSLISIKYLLSHANNFKL
IMFTQWNNFL"
10745..11875
/gene="cag16"
10745..11875
/locus_tag="HP0537"
10745..11875
/gene="cag16"

```

CDS

Query Match	85.3%	Score 584.4	DB 1	Length 26771
Best Local Similarity	94.4%	Pred. No. 1.6e-99		
Matches 606; Conservative	0	Mismatches 36	Indels 0	Gaps 0

Qy	1	AAAAATGGCAAAAATTAAGATTTCCAGCAAGTTAACGCAAGCAAAAAGCAACCTTGAAAT	60
Db	24154	AAAAATGGCAAAAATTAAGATTTCCAGCAAGTTAACGCAAGCAAAAAGCAACCTTGAAAT	24213
Qy	61	TCGGTTAAAGATGTGATCTCATCACTAAAAGTTAACGGTTAAAGTTGATATCTCATCA	120
Db	24214	TCGATTTAAAGATGTGATCTCATCACTAAAAGTTAAAGGATTAAGTTGATCACTCATCA	24273
Qy	121	GGGTTTCAGTGGCTTAAGCAACGGGTGATTTCACTAGGTAGACAGCGTTAGCCGAT	180
Db	24274	GGGTATCAATACCTTAAGCAACGGGTGATTTCACTGGGGGTAGACAGCGCTAGCCGAT	24333
Qy	181	CTCAAAAATTTCTCAAGAGCAATTTGGCCCAACAGCTCAAAAAAATGAAGTCTCAT	240
Db	24334	CTCAAGAAATTTCTCAAAAAGCAATTTGGCTCAACAGCTCAAAAAAATGAAGATTCAT	24393
Qy	241	GCTAGAAAAAATCTGAATATATCATCCGTTAAGAAATGTGTAAATGAAACCTTAGTC	300
Db	24394	ACTGAAAAAATCTGAATATATCAATCCGTTAAGAAATGTGTAAATGAAACCTTAGTC	24453
Qy	301	GGTATGGGTTATCTCAAGCAAGAGCAACACTCTTTCTTAAAACCTTTTGGACATCAAG	360
Db	24454	GGTATGGGTTATCTTAAAGCAAGAGCAACACTCTTTCTTAAAACCTTTTGGACATCAAG	24513
Qy	361	AAAGATGTAATGCCAAAACCTTGAATTTCAATACATTAACATATATGACTCAAAAC	420
Db	24513	AAAGATGTGAGCAAAATTTGGAAATTTCAATACATTAACATATATGACTCAAAAC	24573
Qy	421	GAACCCATTATGCTAAAGTTAATAAAAAAGAAACAGGSCAAGCAGCTAGCTTGAAGA	480
Db	24574	GAACCCATTATGCTAAAGTTAATAAAAAAGAAACAGGCAAGCAGCTAGCCTTGAAGA	24633
Qy	481	CCCATTTAGGCTCAAGTTGCTAAAAGGTAAATGCAAAAATTGACCGACATCAATTA	540
Db	24634	CCCATTTAGGCTCAAGTTGCTTAAAAAGGTAAATGCAAAAATTGACCGACATCAATTA	24693
Qy	541	GCAAGTGTGTTGGGTGTTAGAGGCAACAGCGGGCTCCCTTTGAAAAGCGTATATA	600
Db	24694	GCAAGTGTGTTGGGTGTTAGAGGCAACAGCGGGCTCCCTTTGAAAAGCGTATATA	24753
Qy	601	GTGATGATCTCAATAGATAGAGGCTTTCAAGAAATCAAGAA	642
Db	24754	GTGATGATCTCAATAGATAGAGGCTTTCAAGTAAACATGAA	24795

Search completed: March 6, 2006, 22:39:14
Job time : 3664 secs

Job time : 3664 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:26:07 ; Search time 39 Seconds

(without alignments)
562.498 Million cell updates/sec

Title: US-09-360-685c-25

Perfect score: 1127

Sequence: 1 KNGKNKDFSKYQAKSDLEN.....LSRNOELAQKIDNINQAVSE 228

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1083	96.1	1215	2	B48281
2	959.5	85.1	1186	2	C64588
3	896.5	79.5	1167	2	B71924
4	140	12.4	6713	2	B89921
5	127.5	11.3	774	2	JC2299
6	125	11.1	2481	2	D90011
7	124	11.0	504	2	D71615
8	124	11.0	1302	1	UC6009
9	121	10.7	993	2	E90072
10	121	10.7	1125	2	E90598
11	121	10.7	2401	2	T28676
12	119.5	10.6	821	2	S67087
13	119	10.6	615	2	AB2641
14	118	10.6	622	2	D97423
15	118	10.5	1713	2	A55347
16	117.5	10.4	482	2	C86322
17	117	10.4	672	2	S61463
18	117	10.4	693	2	S61464
19	117	10.4	693	2	I40090
20	116	10.3	667	2	B97012
21	116	10.3	1175	2	D35815
22	116	10.3	1175	2	C35815
23	116	10.3	1201	2	A35815
24	116	10.3	1201	2	B35815
25	116	10.3	2285	2	A32491
26	116	10.3	2411	2	B32491
27	115.5	10.2	1269	2	F84730
28	115	10.2	1191	2	B97116
29	114.5	10.2	481	2	T18465

30	114.5	10.2	1051	2	T18351	lmp1 protein - Myc
31	114.5	10.2	1120	2	A10561	integral membrane
32	114.5	10.2	1365	2	T30822	lmp1 protein - Myc
33	114.5	10.2	1939	1	A67662	myosin alpha heavy
34	114	10.1	1005	2	A64465	hypothetical prote
35	114	10.1	3672	2	T23433	hypothetical prote
36	114	10.1	3704	2	T37316	probable laminin a
37	114	10.1	3712	2	S18253	laminin alpha-1 ch
38	113.5	10.1	722	2	H97217	uncharacterized co
39	113.5	10.1	955	1	A35254	leukotoxin A - pas
40	113.5	10.1	2139	2	T18296	myosin heavy chain
41	113	10.0	1039	2	S62509	probable vesicular
42	113	10.0	1044	2	T50213	probable vesicular
43	113	10.0	1120	2	F90693	mechanosensitive c
44	113	10.0	1120	2	B85544	mechanosensitive c
45	113	10.0	1640	2	D86798	prophage p13 prote

ALIGNMENTS

RESULT 1

B48281

Cytotoxin-associated gene A protein caga - Helicobacter pylori (strains G396 and CUG 1

N.Altredate names: immunodominant 128k antigen; immunodominant 15k antigen

C/Species: Helicobacter pylori

C/Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C/Accession: B48281; A48281

R/Covacci, A.; Censini, S.; Bugnoli, M.; Petracca, R.; Burrone, D.; Macchia, G.; Masso

Proc. Natl. Acad. Sci. U.S.A. 90, 5791-5795, 1993

A/Title: Molecular characterization of the 128-kDa immunodominant antigen of Helicobac

A/Reference number: A48281; MUID:93296225; PMID:8516329

A/Accession: B48281

A/Molecule type: DNA

A/Residues: 1-1215 <COV1>

A/Cross-references: UNIPROT:P80200; UNIPARC:UPI000017A916; GB:X70039; NID:G394912; PIDN

A/Experimental source: strain G396

A/Note: Sequence extracted from NCBI backbone (NCBIN:133971, NCBI:P:133973)

A/Accession type: DNA

A/Residues: 1-957,1026-1215 <COV2>

A/Cross-references: UNIPARC:UPI00002P93F; GB:X70039; NID:G394912; PIDN:CAA49633.1; PID

A/Experimental source: strain CUG 17874

A/Note: Sequence extracted from NCBI backbone (NCBIN:133934, NCBI:P:133968)

C/Comment: These sequences are from cytotoxin producing strains. It is similar to a seq

F:958-1025/Region: 33-residue repeats

Query Match 96.1%; Score 1083; DB 2; Length 1215;

Best Local Similarity 77.0%; Pred. No. 1.8e-57;

Matches 228; Conservative 0; Mismatches 0; Indels 68; Gaps 1;

QY	1	KNGKNKDFSKYQAKSDLENSKVYINOKTIDVDNINQAVSAKTKGDSRREQLAD 60
DB	750	KNGKNKDFSKYQAKSDLENSKVYINQKTDVDNINQAVSAKTKGDSRREQLAD 809
QY	61	LKNSKQELAQQAQKNSINARKSEIYQSVANGVNGTLYGNGLSQAEATTLTKNFSDIK 120
DB	810	LKNSKQELAQQAQKNSINARKSEIYQSVANGVNGTLYGNGLSQAEATTLTKNFSDIK 869
QY	121	KELNAKIGNFNNNNNGKNEPIYAKVKKKGAQASIEEPIYQVAKVNAKIDRLNQi 180
DB	870	KELNAKIGNFNNNNNGKNEPIYAKVKKKGAQASIEEPIYQVAKVNAKIDRLNQi 929
QY	181	ASGLGVGQAAGFPPLKRHDKVDLSKVG 208
DB	930	ASGLGVGQAAGFPPLKRHDKVDLSKVG 989
QY	209	-----LSRNOELAQKIDNINQAVSE 228
DB	990	KVGRSVSPPIYATIDLCGFPPLKRHDKVDLSKVLNRNOELAQKIDNINQAVSE 1045

RESULT 2

C64588
cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)
N/Alternate names: cytotoxin-associated gene A protein
C/Species: Helicobacter pylori
C/Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C/Accession: C64588
R/Tomb, J.F.; White, O.; Kertavage, A.R.; Clayton, R.A.; Sutton, G.G.; Plautsch, R.D.; Peterson, S.; Loftis, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodde, A.; McKenna, N.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Auerhori, S.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A/Reference number: A64520; MIMD:97394467; PMID:9252185
A/Accession: C64588
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 11186 <10M>
A/Cross-reference: UNIPROT:P55980; UNIPARC:UPI000126DD8; GB:AE00569; GB:AE00511, NID

Query Match

Best Local Similarity 85.1%; Score 959.5; DB 2; Length 1186;
Matches 205; Conservative 2; Mismatches 20; Indels 37; Gaps 3;

QY 1 KNGKNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 60
DB 755 KNGKNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 814
QY 61 LKNFSKEQLAQQAQKNEESLNARKKSEIYQSVKNGVNGTIVNGTISOAATTLTKNFSDDIK 120
DB 815 LKNFSKEQLAQQAQKNEEDPVTGKNSLYQSVKNGVNTKIVNGTIGSITATLAKNFSDDIK 874
QY 121 KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 178
DB 875 KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 178
QY 179 QIASGLGVVGOAGPFLKRDHDKVDLSKVG----- 209
DB 934 QIASGLGVVGOAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGPPFLKRDHDKVDLSK 933
QY 210 -----SRNOELAQKIDNLNOAVSE 228
DB 994 SKVGRSRNOELAQKIDNLNOAVSE 1017

RESULT 3

B71924
cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori
C/Species: Helicobacter pylori
A/Variety: strain J99
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C/Accession: B71924
R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Ulang, O.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
A/Reference number: A71800; MIMD:99120557; PMID:9923682
A/Accession: B71924
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1167 <ARN>
A/Cross-reference: UNIPROT:Q9ZUT1; UNIPARC:UPI000126DD8; GB:AE001483; GB:AE001439, NID
A/Experimental source: strain J99
A/Genes: caga

Query Match

Best Local Similarity 79.5%; Score 896.5; DB 2; Length 1167;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 60
DB 754 KNGKNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 813

QY 61 LKNFSKEQLAQQAQKNEESLNARKKSEIYQSVKNGVNGTIVNGTISOAATTLTKNFSDDIK 120
DB 814 LKNFSKEQLAQQAQKNEEDPVTGKNSLYQSVKNGVNTKIVNGTIGSITATLAKNFSDDIK 873
QY 121 KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 180
DB 874 KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 180
QY 181 ASGLGVVGOAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGPPFLKRDHDKVDLSK 206
DB 917 ASGLGVVGOAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGPPFLKRDHDKVDLSK 206
QY 207 VGLSRNOELAQKIDNLNOAVSE 228
DB 977 VGLSRNOELAQKIDNLNOAVSE 998

RESULT 4

B89921
hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C/Accession: B89921
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu, M.; Aizawa, T.; Hattori, M.; Ogata, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MIMD:21311952; PMID:11418146
A/Accession: B89921
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-6713 <KUR>
A/Cross-reference: UNIPROT:Q9U054; UNIPARC:UPI00011021A; GB:BA000018; PID:g13701232, P
A/Experimental source: strain N315
A/Genes: ebha

Query Match 12.4%; Score 140; DB 2; Length 6713;
Best Local Similarity 22.1%; Pred. No. 4.6;
Matches 66; Conservative 50; Mismatches 97; Indels 86; Gaps 11;

QY 5 KNDPSKVTQAKSD-----LENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 54
DB 5825 KNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 5884
QY 5885 KNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 5884
DB 5885 KNDPSKVTQAKSDLENSVDVITINOKYTDKVDNLNOAVSVAKAGDPSRVEQALAD 5884
QY 95 VNGTLVNGTIVNGTISOAATTLTKNFSDDIK-----KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 178
DB 5945 VNGTLVNGTIVNGTISOAATTLTKNFSDDIK-----KEINAKLGNFNNNNNGNLKN--EPYAKVKKKXGAAASLEPIYAVQAKVNAKTIDRLN 178
QY 136 -----NGKNEPIYAVQAKVNAKTIDRLN 178
DB 6004 KQALKDRINQILQGGNDINNNAKTKKEIDQAKQALQALD-----IDQVAKEDKAKNA 6058
QY 179 -----QIASGLGVVGOAGPFLKRDHDKVDLSKVGSLASPEPIYATIDLGPPFLKRDHDKVDLSK 227
DB 6059 IKALANAKRDQINSNPDLTPEQVAKK-----EIDAKKAL--QVENSQTTIDQLNGLN 6113

RESULT 5

UC2299
cell surface glycoprotein MSG99 - Pneumocystis carinii
C/Species: Pneumocystis carinii
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C/Accession: UC2299
R/Wada, M.; Nakamura, Y.; DNA Res. 1, 163-168, 1994
A/Title: MSG gene cluster encoding major cell surface glycoproteins of rat Pneumocystis

F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 11.0%; Score 124; DB 1; Length 1302;
Best Local Similarity 21.0%; Pred. No. 6.3;
Matches 67; Conservative 59; Mismatches 87; Indels 106; Gaps 14;

QY 1 KKGKNDPFGVTOAKSDLENSVQDV--ITNQKTDKVDNLNQAVAKATGDFSRVQAL 58
DB 356 KSNIDNIEKABQSLKSKESMESANDLNTLTLEYEIKLNKQKKEA--KFNLEIQTR 413
QY 59 ADLKNSKEQL-----AQQAKNESLNARKSEIY----- 88
DB 414 KNIENPLTDEVKNNPNVATLVKDLTNAKDAKKSVT--NSSKSDIITANBALIQALDANK 472
QY 89 -----QSVKNGVNGTLVNGLSQAEATTLKNSFD-----IKKELNATLGNFN- 131
DB 473 AKQVDENAKSIKEQLNALI-----DKAVTLLPQLNDNSEIYKAKESLNMEITNAK 525
QY 132 --NNNNNG-----LKNPIYAKVKKKAGQAASLE----- 160
DB 526 AVNQNDNAMSQAKSLDVKYTKIQNQ--LTFPKQDAKFKELQTKKIDNPLTDDYK 583
QY 161 --PIYAOVAKKV--NAKIDRLNQAISGLGVGAAGPPIK-----HDKVDLSKVG 209
DB 584 NNPNVATLVKDLTNAKDKKSVTKSNKSEIIANDELQALDRAKAKAQQDIDENAK-- 640
QY 210 SRNOELAKIDNLNQAVSE 228
DB 641 SIKEQLSDSITNANQILNK 659

RESULT 9

hypothetical protein SA2436 [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C90072
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, A.; Mizutani-Ju, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogatawara, N.; Hayashi, H.; Hiramatsu, K.
L:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A69758; MUID:21311952; PMID:11418146
A:Accession: C90072
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-993 <KUR>
A:Cross-references: UNIPROT:Q99025; UNIPARC:UPI00000CAB7A; GB:BA000018; PTD:913702601; E:Experimental source: strain N315
C:Genetics:
A:Gene: SA2436

Query Match 10.7%; Score 121; DB 2; Length 993;
Best Local Similarity 25.0%; Pred. No. 7;
Matches 66; Conservative 45; Mismatches 79; Indels 74; Gaps 16;

QY 12 TOAKS-----DIENSVQDVITNQKTDKVDNLNQAVAKATGDFSRVQALDANKS 65
DB 457 TOAKAQQKDSQALRNISYGIASDKPSDFRSLDVGSGLEFYTQYN--QQPITLLEIE 514
QY 66 KEQ---LAQQAOK-----NESLNARKSEIYQSVKNGVNGTLVNGLSQAEAT-- 110
DB 515 KNEVDLSKEIDKYAANNRNESL--RLVQLSVALNASSST-----AAATLTD 564
QY 111 TLK-----NPSD-IKKEINATLGNFN-----NNNNGLKNEPIYAKVKK--KTA 152
DB 565 QLSKIDSLSSFDVYKDDLNSSLSVISKRIWDELNKGQTALSN--VOSKNTIDIDVINS 622
QY 153 GQAASLEPIYAOVAKKVAKIDRLNQAISGL-----GVGQAAGPPLKHDKVDLSK 206
DB 623 GQA-----ITKNGKTRIDRLQTLVPSIEQOYISAVKNAQANFPKYSVDVAAANF 672
QY 207 VGLSRNQ--ELAKIDNLNQAVSE 228

DB 673 V--RNDLPQBLQRLTNATVAVNK 693

RESULT 10

membrane nuclease, lipoprotein [imported] - *Mycoplasma pulmonis* (strain UNB CTTP)
C:Species: *Mycoplasma pulmonis*
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: E90598
R:Chambaud, I.; Halli, R.; Ferris, S.; Barde, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*.
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: E90598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1125 <KUR>
A:Cross-references: UNIPROT:Q99PM9; UNIPARC:UPI00000D464B; GB:AL445566; PTD:914090108; E:Experimental source: strain UNB CTTP
C:Genetics:
A:Gene: MYPV 6930
A:Genetic code: SGC3

Query Match 10.7%; Score 121; DB 2; Length 1125;
Best Local Similarity 24.1%; Pred. No. 8.1;
Matches 66; Conservative 41; Mismatches 107; Indels 60; Gaps 13;

QY 2 NGKND-----PSKYTOAKSDLENSVQDVITNQKTDKVDN-----NOAVSVAKAT 48
DB 581 NGENDSKQNTSNRQTNDLSEKQNLTKKPSNNSNVETKQETONNENSTKDE 640
QY 49 GDSRVEQ--ALADLKNSKEQLAQAQAKNESLNAR-----KKEIYQSVKNGV 95
DB 641 IDISACTQDSTNSLNKKEKTQVTKTNTSNNSNSTKQENSTKKEISKESN-V 699
QY 96 NGTLVNGLSQAEATTLKNSFDIKKELNATLGNFNNNNNNLKNEPIYAKVKK----- 150
DB 700 NNS--NSTNQE-----ENIINKKEEISKESNNSNSTVQNET--PETESQNNVI 750
QY 151 --KQAASLEPIYAOVAKKVAKIDRLNQAISGLGVGAAGPPIK-----HDKVDLS 205
DB 751 IGKPNMQSLNQANIDVSAKV--KIGYMN--INESVGSASAKPAKVIDHNLDLVG 807
QY 206 KVGLSRNOELAKIDNLN-----QAVSE 228
DB 808 IGGVHEETLTKIVEEMNLSKDSQDKWQVISE 841

RESULT 11

rhoptry protein - *Plasmodium yoelii* (fragment)
C:Species: *Plasmodium yoelii*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
R:Simha, K.A.; Keen, J.K.; Ogum, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a *Plasmodium yoelii* rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <81N>
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:91041784; PI
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a *Plasmodium yoelii* rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: T28676
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281


```
QY      76 NESLNARKKSEIYOSKGVN-----GTLVNGLSQAETTLTKNPFSDIKKELNKLGN 129
Db      293 ERHNDARKQELDQIEPAGVEIASGLGRSLRSLRTETPACRLDRLRTPNESLIN 352
QY      130 FNNN-----NNNGLANEPIYAKVKKKAGQAASLEPIYA--QVAKVNAKIDR 176
Db      353 LRDALGOIREFTLLIQNSGIEIQQSSVDLSKRTENQAASLEETAAVAIEITATVRSASR 412
QY      177 LNOIASGLGVGOAA---GPELKHDKYDDLSKY-GLSRN-DELAOKIDNL 222
Db      413 AREANEAVATVKOSADSSGSVGS--NAVDAKSRIGASRKIEQIIEVYDDI 461
```

RESULT 15

```
AS5347
adhesive ligand ep11gtrn, alpha-3 chain form A precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Feb-1995 #sequence_revisions 06-Feb-1995 #text_change 09-Jul-2004
C:Accession: AS5347
R:Ryan, W.C.; Tizard, R.; Vandevanter, D.R.; Carter, W.G.
J:BioJ. Chem. 269, 22779-22787, 1994
A:Title: Cloning of the lamA3 gene encoding the alpha3 chain of the adhesive ligand ep1
A:Reference number: AS5347; WUID:94357926; PMID:8077230
A:Accession: AS5347
A:Status: P:155347
A:Molecule type: primary
A:Residues: 15713 <RA>
A:Cross-reference: UNIPROT:Q16787; UNIPARC:UPI00000349C3; GB:L34155; NID:9551596; PID:9
C:Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like
F:67-114/Domain: laminin-type EGF-like homology <LE2>
F:1192-1534/Domain: laminin G repeat homology <LG4>
```

Query Match 10 5% Score 118; DB 2; Length 1713;
Best Local Similarity 21.28; Pred. No. 20; Mismatches 107; Indels 50; Gaps 9;

```
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;
QY      5 NKDFSKVTQAKSDLENSKDVIIINOKVTQVNDLNOAVSAKATGDFSR---VEQALAD 60
Db      300 NNNVNRATQSAKEIDVIRIVINVAHILKQISGTDGEGNNVPSGDFSRMAEAQRMRE 359
QY      61 L--KNFSKEQLAQQAQKNEISLNARKKSEIYOSKGVNNGTLVNGLSQAETTLTK--N 115
Db      360 LNRNRFQGLREARADKRSSQLLRIRTKWOKTHQSEN-----NGLANSIRDSLNEYEAK 414
QY      116 PSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVKK-----KAGQAASLEE 160
Db      415 LSDLRARLQEAQAQAQANGLNQENBERALGATQGVKEINSLSGDFTKYITTADESLIQ 474
QY      161 PIYAQVAKVNAKIDRLNIOIASGLGVGOAAAGPLKRDHYDDLS---KVGLSRN--- 212
Db      475 NIALQLMEKSKQKEYEL---AASINERQ-----ELSDKYRELRSAGKTSLVVEAKH 525
QY      213 ---DELAOKIDNLNOAVS 227
Db      526 ARSLQELAKQLEIRIKNAS 544
```

Search completed: March 6, 2006, 20:30:33
Job time : 40 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:41:18 ; Search time 163 Seconds

(without alignments)
584,448 Million cell updates/sec

Title: US-09-360-685c-25

Perfect score: 1127
Sequence: 1 KNGKNKDPSTQAKSDLEN.....LSRNGELAKIDNNGAYSE 228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1127	100.0	1147	3	US-09-921-157-5 Sequence 5, Appl1
2	1127	100.0	1138	4	US-09-402-100-4 Sequence 4, Appl1
3	896.5	79.5	1167	4	US-10-335-977-8633 Sequence 8633, Ap
4	896.5	79.5	1178	4	US-10-335-977-8634 Sequence 8634, Ap
5	896.5	79.5	1183	4	US-10-335-977-8635 Sequence 8635, Ap
6	140	12.4	6713	4	US-10-282-122A-43811 Sequence 43811, A
7	131.5	11.7	1404	5	US-10-732-923-3304 Sequence 3304, A
8	131	11.6	3533	4	US-10-282-122A-70177 Sequence 70177, A
9	129	11.4	1992	5	US-10-470-048B-410 Sequence 81, Appl
10	129	11.4	10498	5	US-10-470-048B-410 Sequence 440, Appl
11	127	11.3	837	3	US-09-815-242-5883 Sequence 5883, Ap
12	127	11.3	875	3	US-09-815-242-13080 Sequence 13080, A
13	127	11.3	1361	4	US-10-369-493-3209 Sequence 3209, Ap
14	127	11.3	2434	3	US-09-815-242-5835 Sequence 5835, Ap
15	127	11.3	6281	3	US-09-815-242-12996 Sequence 12996, A
16	126	11.2	2368	3	US-09-815-242-5635 Sequence 5635, Ap
17	126	11.2	2368	3	US-09-815-242-12389 Sequence 12389, A
18	125.5	11.1	2437	3	US-09-815-242-5834 Sequence 5834, Ap
19	125	11.1	2478	3	US-09-815-242-5816 Sequence 5816, Ap
20	125	11.1	2478	3	US-09-815-242-12967 Sequence 12967, A
21	125	11.1	2478	3	US-10-470-048B-320 Sequence 220, Appl
22	125	11.1	2481	4	US-10-282-122A-43762 Sequence 43762, A
23	124	11.0	5795	3	US-09-815-242-12610 Sequence 12610, A
24	123	10.9	10203	4	US-10-661-809-23 Sequence 23, Appl
25	123	10.9	10203	4	US-10-724-972A-4098 Sequence 4098, Ap
26	121.5	10.8	6641	4	US-10-282-122A-70580 Sequence 70580, A
27	121	10.7	993	4	US-10-282-122A-43875 Sequence 43875, A

28	120	10.6	3158	3	US-09-815-242-12611 Sequence 12611, A
29	119	10.6	998	4	US-10-282-122A-70450 Sequence 70450, A
30	118	10.5	724	4	US-10-282-122A-71401 Sequence 71401, A
31	118	10.5	1441	5	US-10-732-923-3352 Sequence 3352, Ap
32	118	10.5	1693	4	US-10-603-725-4 Sequence 4, Appl1
33	118	10.5	1693	4	US-10-603-725-8 Sequence 8, Appl1
34	118	10.5	1713	4	US-10-171-311-113 Sequence 113, Appl
35	118	10.5	1713	4	US-10-372-683-10 Sequence 10, Appl
36	118	10.5	1713	4	US-10-603-725-6 Sequence 6, Appl1
37	118	10.5	1724	4	US-10-603-725-2 Sequence 2, Appl1
38	117.5	10.4	1047	4	US-10-282-122A-44353 Sequence 44353, A
39	117.5	10.4	2025	3	US-09-815-242-5703 Sequence 5703, Ap
40	117	10.4	693	4	US-10-369-100-68 Sequence 68, Appl
41	116	10.3	2067	6	US-11-097-143-40167 Sequence 40167, A
42	116	10.3	2539	5	US-10-831-070-6 Sequence 6, Appl1
43	115.5	10.2	1090	4	US-10-282-122A-56217 Sequence 56217, A
44	115	10.2	873	3	US-09-952-267-13 Sequence 13, Appl
45	115	10.2	873	5	US-10-872-766-13 Sequence 13, Appl

ALIGNMENTS

```
RESULT 1
US-09-921-157-5
Sequence 5, Application US/09921157
Publication No. US20040048353A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonio
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Cytotoxin Proteins Useful For
FILE REFERENCE: CHIR0315
CURRENT APPLICATION NUMBER: US/09/921,157
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/466,662
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,846
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/360,934
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: PCT/EP93/00472
PRIOR FILING DATE: 1993-03-02
PRIOR APPLICATION NUMBER: PCT/EP93/00158
PRIOR FILING DATE: 1993-01-25
PRIOR APPLICATION NUMBER: FI 92 A 000052
PRIOR FILING DATE: 1992-03-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-921-157-5
Query Match 100.0%; Score 1127; DB 3; Length 1147;
Best Local Similarity 100.0%; Pred. No. 9,7e-78;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 KNGKNKDPSTQAKSDLENSKVITINOKTIDVDINLQAVSAKATGDSRYEQALAD 60
|||||
750 KNGKNKDPSTQAKSDLENSKVITINOKTIDVDINLQAVSAKATGDSRYEQALAD 809
|||||
61 LKNSKEQLAQQAQKNSISNARKSEIYQSVKGVNGLVNGSGQAEATLTSKNFSDIK 120
|||||
810 LKNSKEQLAQQAQKNSISNARKSEIYQSVKGVNGLVNGSGQAEATLTSKNFSDIK 869
|||||
121 KELNAKLGNNNNNNNGIKNEPIYAKVKKKAGQAASLEBPIYAQVAKKVAKIRLNQI 180
```

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rapbm

Page 2

Db 870 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 228
Db 930 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 977

RESULT 2

US-09-402-100-4
Sequence 4, Application US/09402100
Patent No. US20010019834A1
GENERAL INFORMATION:
APPLICANT: Daewoong Pharmaceutical Co, LTD
APPLICANT: Kim, Byung-O
APPLICANT: Shin, Sung-Seup
APPLICANT: Yu, Young-Hyo
APPLICANT: Park, Myung-Hwan
APPLICANT: Choi, Deok-Joon
APPLICANT: Jung, Hyung-Jin
TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Helicobacter pylori
FILE REFERENCE: 0136/06140
CURRENT APPLICATION NUMBER: US/09/402,100
EARLIER APPLICATION NUMBER: KR 97-11950
EARLIER FILING DATE: 1997-03-31
EARLIER APPLICATION NUMBER: KR 97-11951
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 1338
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: GAGA/CTAA2B Chimeric protein
US-09-402-100-4

Query Match
Best Local Similarity 100.0%; Score 1127; DB 3; Length 1338;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKKKKKDSKYTKQAKSDLENSVYDIINOKVTDKVDNLMQAVAKATGDFSRVEQALAD 60
Db 750 KKKKKKDSKYTKQAKSDLENSVYDIINOKVTDKVDNLMQAVAKATGDFSRVEQALAD 809
QY 61 LKNSFKEQLAQQAQKNSLNAKKESEIYOSVKGNGVGLVNGLSQAEKTTLSKNFSDIK 120
Db 810 LKNSFKEQLAQQAQKNSLNAKKESEIYOSVKGNGVGLVNGLSQAEKTTLSKNFSDIK 869
QY 121 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 180
Db 870 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 929
QY 181 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 228
Db 930 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 977

RESULT 3

US-10-335-977-8633
Sequence 8633, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS: 10031
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston

STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8633:
SEQUENCE CHARACTERISTICS:
LENGTH: 1167 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1167
SEQUENCE DESCRIPTION: SEQ ID NO: 8633:
US-10-335-977-8633

Query Match
Best Local Similarity 79.5%; Score 996.5; DB 4; Length 1167;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KKKKKKDSKYTKQAKSDLENSVYDIINOKVTDKVDNLMQAVAKATGDFSRVEQALAD 60
Db 754 KKKKKKDSKYTKQAKSDLENSVYDIINOKVTDKVDNLMQAVAKATGDFSRVEQALAD 813
QY 61 LKNSFKEQLAQQAQKNSLNAKKESEIYOSVKGNGVGLVNGLSQAEKTTLSKNFSDIK 120
Db 814 LKNSFKEQLAQQAQKNSLNAKKESEIYOSVKGNGVGLVNGLSQAEKTTLSKNFSDIK 873
QY 121 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 180
Db 874 KEINAKLGNFNNNNNGLKNEPIYAKVKKGAQAASLEPIYAQVAKVNAKIDRLNOI 916
QY 181 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 228
Db 917 ASGLGVVGOAAGFPLKRDHVDLSTVGLSNRQELAKIDINLNOAVSE 977
QY 207 VGLSRQELAKIDINLNOAVSE 228
Db 977 VGLSRQELAKIDINLNOAVSE 998

RESULT 4

US-10-335-977-8634
Sequence 8634, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS: 10031
ADDRESS: LAHIVE & COCKFIELD

STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7480
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8634:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1178
SEQUENCE DESCRIPTION: SEQ ID NO: 8634:
US-10-335-977-8634

Query Match 79.5%; Score 896.5; DB 4; Length 1178;
Best Local Similarity 72.1%; Pred. No. 5e-60;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKVTQAKSDLENSVQVITNOKVTDKVDNINQAVSVAKATGDFSRVQALAD 60
DB 765 KNGKNDPSKVTQAKSDLENSVQVITNOKVTDKVDNINQAVSVAKATGDFSRVQALAD 824
QY 61 LKNFSKEQLAQQAQKNEISINARKSEIYOSVKNVNGTIVGNGLSQAERTTLISKNFSDIK 120
DB 825 LKNFSKEQLAQQAQKNEIDFNTGKNSALYOSVKNVNGTIVGNGLSQAERTTLISKNFSDIK 884
QY 121 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAQVAKKVNAKIDRLNOI 180
DB 885 KEINAKLGPNPNNNNNGLENS-----TEPIYQVAKKVNAKIDRLDOI 927
QY 181 ASGLGVVGOAG-----PFLKRHDKYDLSK 206
DB 928 ASGLGVVGOAGSFLKRHDKYDLSKVLGSANHEPIYATIDLGPFPLKRHDKYDLSK 987
QY 207 VGLSRNQLAQKIDNINQAVSE 228
DB 988 VGLSRNQLAQKIDNINQAVSE 1009

RESULT 5
US-10-335-977-8635
Sequence 8635, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 8635:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...1183.
SEQUENCE DESCRIPTION: SEQ ID NO: 8635:
US-10-335-977-8635

Query Match 79.5%; Score 896.5; DB 4; Length 1183;
Best Local Similarity 72.1%; Pred. No. 5e-60;
Matches 189; Conservative 7; Mismatches 15; Indels 51; Gaps 2;

QY 1 KNGKNDPSKVTQAKSDLENSVQVITNOKVTDKVDNINQAVSVAKATGDFSRVQALAD 60
DB 770 KNGKNDPSKVTQAKSDLENSVQVITNOKVTDKVDNINQAVSVAKATGDFSRVQALAD 829
QY 61 LKNFSKEQLAQQAQKNEISINARKSEIYOSVKNVNGTIVGNGLSQAERTTLISKNFSDIK 120
DB 830 LKNFSKEQLAQQAQKNEIDFNTGKNSALYOSVKNVNGTIVGNGLSQAERTTLISKNFSDIK 889
QY 121 KEINAKLGPNPNNNNNGLKNEPIYAKVNKKKAGQAASLEPIYAQVAKKVNAKIDRLNOI 180
DB 890 KEINAKLGPNPNNNNNGLENS-----TEPIYQVAKKVNAKIDRLDOI 932
QY 181 ASGLGVVGOAG-----PFLKRHDKYDLSK 206
DB 933 ASGLGVVGOAGSFLKRHDKYDLSKVLGSANHEPIYATIDLGPFPLKRHDKYDLSK 992
QY 207 VGLSRNQLAQKIDNINQAVSE 228
DB 993 VGLSRNQLAQKIDNINQAVSE 1014

RESULT 6
US-10-282-122A-43811
Sequence 43811, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rep.bm

Page 4

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43811
; LENGTH: 6713
; TYPE: FRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-43811

```

```

Query Match 12.4%; Score 140; DB 4; Length 6713;
Best Local Similarity 22.1%; Pred. No. 0.54;
Matches 66; Conservative 50; Mismatches 97; Indels 86; Gaps 11;

QY 5 NKDFSKVTQAKSD-----LENSYKDV---INOKVTDKVDNINQAVSVAKTGDPSRV 54
DB 5825 NNDVVKQVQALIDEIRNPNLTDKEKQALDKRINQILQGHNDINNALTKEEIEQAKQL 5884
QY 55 EQALADLKNFSK-----EQLAQQAQKESINARKK-----SEIYQSVANG 94
DB 5885 AQAQDIDIKLVAKEDAKQDVQVQALIDEIDONPNLTDKEKQALDKRINQILQGHNG 5944
QY 95 VNGTLVNGSLSGAEATLTSKNFSDIK-----KELNAKLGNFNNNNN----- 135
DB 5945 INNAMTKEEIEQAK--QLAQALKEIKLVYAKENAKQDVQVQALIDEIDONPNLTDKE 6003
QY 136 -----NGLKNEPIYAKVKKKGAQGAASLEPIYAKVAKVAKIDRLN- 178
DB 6004 KQALKRINQILQGHNDINNAMTKEEIEQAKQALQALD-----IKLVYAKEDAKVA 6058
QY 179 -----QIASLGVVGAAGFPLKXHDVDDLSTVGLSHNDELQKIDNINQAVS 227
DB 6059 IKALNAKRDQINSNPDLTPQAKALK--EIDBAEKRAL-QVNNMOTIDQNGRLN 6113

```

RESULT 7
US-10-732-923-3304
; Sequence 3304, Application US/10732923
; Publication No. US2005010879A1
; GENERAL INFORMATION:
; APPLICANT: Edgerston, Michael D

```

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(527)96C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3304
; LENGTH: 1404
; TYPE: FRT
; ORGANISM: Mycoplasma hominis
US-10-732-923-3304

```

```

Query Match 11.7%; Score 131.5; DB 5; Length 1404;
Best Local Similarity 20.6%; Pred. No. 0.34;
Matches 51; Conservative 60; Mismatches 97; Indels 39; Gaps 9;

QY 1 KXGKXKDFSKVTQAKSDLENSYKDV---INOKVTDKVDNINQAVSVAKT--GDFSRVQ 57
DB 112 KXKXKHEBDQIKIVQAMQEFKKSQKALGDLINSDGGRVNSAKOSIQNTNNNSISIQ 171
QY 58 LADLKNFSKQALQQAQKNEISINARKKSEIYQSVANGNGLVNGL---SQAETTLK 114
DB 172 IQALSKINEAKKELOSQINNARNQEK--EVEFEKQQLNLINSNEDNSKKADETAIK 229
QY 115 NFSDI-----KELNAKLGNFNNNNNNGLKNEPIYAKVKKKGAQGAASLEPIY 163
DB 230 NTNVAVDSDIKITFKTKKEIEKALISLTNKLNE--FKGQKAWAVASFKSKQALDLD 268
QY 164 AQVAKVNAKIDRLNQIASLGVVGAAGFPLKXHDVDDLSTVGL--LSRNOBLAQKDN 221
DB 289 SEDCKVQSSNE--SQVLT-----KTKIDNSSIDIDNKTQIEKAIKS 331
QY 222 LNQAVSE 228
DB 332 LTNKIND 338

```

```

RESULT 8
US-10-282-122A-70177
; Sequence 70177, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining prior application data removed - See file wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 70177
 LENGTH: 3533
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-10-282-122A-70177

Query Match 11.6%; Score 131; DB 4; Length 3533;
 Best Local Similarity 22.0%; Pred. No. 1.2;
 Matches 66; Conservative 46; Mismatches 100; Indels 88; Gaps 11;

QY 5 NDFSKVTOAKSD-----LENSVVDV---INQVTDKVDNLQAVAKATGDFSKV 54
 DB 2645 NNDVQKVOALIDEIDRNPLTDKQKALDRINQLOOGHGINNMTKEIEQAKQL 2704
 QY 55 EQALADLNKFSK-----EQLAQQAQKESLNARKK-----SEIYQSVKNG 94
 DB 2705 AQAQDIDKLVAKEDAKQDVQKQVALIDEIDQNPNTDKQKALDRINQLOOGHND 2764
 QY 95 VNGTLVNGLSQA-----EATLSKPSFIIKELNAKGNFNNNNN--- 135
 DB 2765 INNMTKEIEQAKERLAQALDIDKLVAKEDAKK--DIDKRVQALIDEIDQNPNTDK 2822
 QY 136 -----NGLKNEPIYAKVKKKKAGAAASLEPIYQVAKVNAKIDRLN 178
 DB 2823 EKQALKDRINQLOOGHNDINNALTKEIEQAAQALQALQD-----IDLVAKEDAKN 2877
 QY 179 -----QIASGLGVVQAAGFPLRKHDKVDLSKVGSRNOELAQKIDNINQAVS 227
 DB 2878 AIKALANAKRDQINSNPDLTPEQKAKALK--EIDBAERKAL-QVVENAQITDQINRGLN 2933

RESULT 9
 US-10-470-048B-81
 Sequence 81, Application US/10470048B
 Publication No. US20050037444A1
 GENERAL INFORMATION:
 APPLICANT: MEINKE ET AL.
 TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
 FILE REFERENCE: SONN.035US
 CURRENT APPLICATION NUMBER: US/10/470,048B
 NUMBER OF SEQ ID NOS: 603
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 81
 LENGTH: 1992
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: MOD RES
 LOCATION: (1222)..(1225)
 OTHER INFORMATION: X = anything
 US-10-470-048B-81

Query Match 11.4%; Score 129; DB 5; Length 1992;
 Best Local Similarity 24.3%; Pred. No. 0.81;
 Matches 65; Conservative 40; Mismatches 71; Indels 92; Gaps 13;
 QY 6 KDFSKVTO-AKSDLENSVVDV---INQVTDK-----VNLNQ----- 40
 DB 1171 KDLVAKEDAKQDVQKQVALIDEIDQNPNTDKQKALDRINQLOOGHNDIXNMTK 1230
 QY 41 -AVSAKATGDFSRVQALADLNKFSK-----EQLAQQAQKESLNARKK-- 84

DB 1231 EALBQAK-----ERLAQALQIDKLVAKEDAKNDIDKRVQALIDEIDQNPNTDKQKA 1285
 QY 85 -----SEIYQSVKNGVNGTLVNGLSQAQATLSKPSFIIKELNAKGNFNNNNNGIK 139
 DB 1286 LKDRINQLOOGHNDINNALTKEIEQAKA-QLAQALQIK----- 1325
 QY 140 NEPIYAKVKKKAGAAASLEPIYQVAKVNAKIDRLNQIASGLGVVQAAGFPLKRD 199
 DB 1326 -DLVAKEDAKNAIKALA-----NAKRDQIN--SNPDLTPEQKAKALK-- 1365
 QY 200 KYDDLSTKVGSRNOELAQKIDNINQAVS 227
 DB 1366 EIDBAERKAL-QVVENAQITDQINRGLN 1392

RESULT 10
 US-10-470-048B-440
 Sequence 440, Application US/10470048B
 Publication No. US20050037444A1
 GENERAL INFORMATION:
 APPLICANT: MEINKE ET AL.
 TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
 FILE REFERENCE: SONN.035US
 CURRENT APPLICATION NUMBER: US/10/470,048B
 NUMBER OF SEQ ID NOS: 603
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 440
 LENGTH: 10498
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 FEATURE:
 NAME/KEY: MOD RES
 LOCATION: (9728)..(9731)
 OTHER INFORMATION: X = anything
 US-10-470-048B-440

Query Match 11.4%; Score 129; DB 5; Length 10498;
 Best Local Similarity 24.3%; Pred. No. 6.6;
 Matches 65; Conservative 40; Mismatches 71; Indels 92; Gaps 13;

QY 6 KDFSKVTO-AKSDLENSVVDV---INQVTDK-----VNLNQ----- 40
 DB 9677 KDLVAKEDAKQDVQKQVALIDEIDQNPNTDKQKALDRINQLOOGHNDIXNMTK 9736
 QY 41 -AVSAKATGDFSRVQALADLNKFSK-----EQLAQQAQKESLNARKK-- 84
 DB 9737 EALBQAK-----ERLAQALQIDKLVAKEDAKNDIDKRVQALIDEIDQNPNTDKQKA 9791
 QY 85 -----SEIYQSVKNGVNGTLVNGLSQAQATLSKPSFIIKELNAKGNFNNNNNGIK 139
 DB 9792 LKDRINQLOOGHNDINNALTKEIEQAKA-QLAQALQIK----- 9831
 QY 140 NEPIYAKVKKKAGAAASLEPIYQVAKVNAKIDRLNQIASGLGVVQAAGFPLKRD 199
 DB 9832 -DLVAKEDAKNAIKALA-----NAKRDQIN--SNPDLTPEQKAKALK-- 9871
 QY 200 KYDDLSTKVGSRNOELAQKIDNINQAVS 227
 DB 9872 EIDBAERKAL-QVVENAQITDQINRGLN 9898

RESULT 11
 US-09-815-242-5883
 Sequence 5883, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zvekind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.

```
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA 011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2001-02-16
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 583
/ LENGTH: 837
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-5883

Query Match
Best Local Similarity 11.3%; Score 127; DB 3; Length 837;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNKDSKYTOAKSDLENSVQDVI---NOKVTD---KVDMLNOAVSVA----- 45
DB 470 GNEOKANNVDAMDKLQSIADNATTKONQVYTDASQNKDYNNAVTTAGIIDDITSP 529
QY 46 -----KATGDSRYEQLADLNKF--SKEQLAQQAQKNSLNARKSEIYQSVKNGV 95
DB 530 TLDPTVYINQAQGVSTTKNALNGNENLEAAQQAQSSLSLSDNLNNAOK---QVTDOI 585
QY 96 NG-----TLVGN---GLSQAEATTLISKPSDIKKEELNAKLGNNNN 134
DB 586 NGAHTVDEANQIKONQNTMTMGKLAIDADATKATVFTDADQ---AKQAQVNTAV 642
QY 135 NNGELNKEPIYAKVKKKGAQAASLEPIYQVAKKYNKIDRLNQASLGIVGQAAGFP 194
DB 643 TNA---ENIISKANGNATQAE-----VQAIKQVNAKAQALNGANN----- 681
QY 195 LKRHDKVDLSKVGLSRNOELAQKIDMLNOAV 226
DB 682 -VOHAKDEATVTLINSSNDLNOAK-DALKQOV 711

RESULT 12
US-09-815-242-13080
/ Sequence 13080, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlson, Karl L.
/ APPLICANT: Eyskens, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ TITLE OF INVENTION: Ident
/ FILE REFERENCE: ELITRA 011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
```

```
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13080
/ LENGTH: 875
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-13080
```

```
Query Match
Best Local Similarity 11.3%; Score 127; DB 3; Length 875;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNKDSKYTOAKSDLENSVQDVI---NOKVTD---KVDMLNOAVSVA----- 45
DB 507 GNEOKANNVDAMDKLQSIADNATTKONQVYTDASQNKDYNNAVTTAGIIDDITSP 566
QY 46 -----KATGDSRYEQLADLNKF--SKEQLAQQAQKNSLNARKSEIYQSVKNGV 95
DB 567 TLDPTVYINQAQGVSTTKNALNGNENLEAAQQAQSSLSLSDNLNNAOK---QVTDOI 622
QY 96 NG-----TLVGN---GLSQAEATTLISKPSDIKKEELNAKLGNNNN 134
DB 623 NGAHTVDEANQIKONQNTMTMGKLAIDADATKATVFTDADQ---AKQAQVNTAV 679
QY 135 NNGELNKEPIYAKVKKKGAQAASLEPIYQVAKKYNKIDRLNQASLGIVGQAAGFP 194
DB 680 TNA---ENIISKANGNATQAE-----VQAIKQVNAKAQALNGANN----- 718
QY 195 LKRHDKVDLSKVGLSRNOELAQKIDMLNOAV 226
DB 719 -VOHAKDEATVTLINSSNDLNOAK-DALKQOV 748
```

```
RESULT 13
US-10-369-493-3209
/ Sequence 3209, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 3209
/ LENGTH: 1361
/ TYPE: PRT
/ ORGANISM: Neurospora crassa
US-10-369-493-3209
```

```
Query Match
Best Local Similarity 11.3%; Score 127; DB 4; Length 1361;
Matches 24.4%; Pred. No. 0.72;
```


Matches 76; Conservative 40; Mismatches 92; Indels 104; Gaps 14;

QY 6 KDFSKVTQAKSD-LENSVQV-----IINOK-----VTRKDNINQAV 42
DB 1003 KDLTKQSQTREADLNDNDIKMKSTKEKLGQDAPAKETVLAERKEIIGQLDSDINRLNODI 1062
QY 43 SVAKATGDSR--VEQALADLK-----NFSKEQLAQOAO-----74
DB 1063 STGNATLDKREIIDDLDKIDKIANSTIDTLKRDVADKQALILAKTKDVAARPAELAK 1122
QY 75 -----KNEELNARKSEIYOSVNGVNGTLVNGLSQAEAT-----TLSKNPSDI 119
DB 1123 ARIASKNAL-AKTEBAKAFERKNVQTLTDQAKGLNDVATKTYQLAQBRAITISKLNKOI 1181
QY 120 -----KKEINAKLGNF-----NNNNNGK--NEPIYAK--VAKKAGQASIE 159
DB 1182 FDLKTDVTKKQSLSTKDNALTKQAGEISGRDGLAKREBELAKKALAKKTEERASSIE 1241
QY 160 EPIYAQVAKVNAKID--RLNQIASGLGVGOAGFPLKRBKVDLSKVGLSRQELA 216
DB 1242 KVKKLTDEATGAKDVTSTRTQLA-----ODKDAISKLEKDIKAKLNQELS 1287
QY 217 QKIDNINQAVSE 228
DB 1288 TKDASLTQKTGE 1299

RESULT 14
US-09-815-242-5835
Sequence 5835, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 5835
LENGTH: 2434
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5835

Query Match 11.3%; Score 127; DB 3; Length 2434;
Best Local Similarity 25.4%; Pred. No. 1.5;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNDPSKVTQAKSDLENSVKDVI--NOKVTD--KVDNINQAVSA-----45

DB 2130 GNEQKANNVDAAMDKRQSIADNATTKQNNTTDSQNKKAYNNAVTTAQSIIDQTTSP 2189
QY 46 -----KATGDSRVEQALADLKNF--SKEQLAQOAKNEELNARKKSEIYOSVNGV 95
DB 2190 TLDPFTVINQAGVSTTKNALNGNENLEAKQOASQSLGSLDNLNNAOK--QVTVDQI 2245
QY 96 NG-----TLVGN--GLSQAEATTLKSNPSDIKELNARKLGNFNNN 134
DB 2246 NGAHTVDEANQIKQNAQNLNTMGNLKOALIDKDATKATVNTFDADQ--AKQOAYNTAV 2302
QY 135 NNGELNKEPIYAKVKKKAGQASLREPIYAQVAKVNAKIDRLNQIASGLGVGOAGF 194
DB 2303 TNA--ENIISAKANGATQAE-----VEQAIKQVNAKAKALNGNAN-----2341
QY 195 LKRHDKVDLSKVGLSRQELAQKIDNINQAV 226
DB 2342 -VOHAKDEATLALINSNDLNQAK--DALKQOV 2371

RESULT 15
US-09-815-242-12996
Sequence 12996, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 12996
LENGTH: 6281
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 11.3%; Score 127; DB 3; Length 6281;
Best Local Similarity 25.4%; Pred. No. 4.9;
Matches 69; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

QY 3 GKNDPSKVTQAKSDLENSVKDVI--NOKVTD--KVDNINQAVSA-----45
DB 2909 GNEQKANNVDAAMDKRQSIADNATTKQNNTTDSQNKKAYNNAVTTAQSIIDQTTSP 2968
QY 46 -----KATGDSRVEQALADLKNF--SKEQLAQOAKNEELNARKKSEIYOSVNGV 95
DB 2969 TLDPFTVINQAGVSTTKNALNGNENLEAKQOASQSLGSLDNLNNAOK--QVTVDQI 3024
QY 96 NG-----TLVGN--GLSQAEATTLKSNPSDIKELNARKLGNFNNN 134

Tue Mar 7 12:58:23 2006

us-09-360-685c-25.rapbm

Page 8

Db 3025 NGAHTVDEANOIKONQNLNTAMGNLQALADKDATNATVFTDADQ--AKQOAYNTAV 3081
QY 135 NNGIKNEPIYAKVKKKXAGQASLEPIYAOVAKKVNAKIDRLNOIASGLGVVGOAAGFP 194
Db 3082 TNA---ENIISKANGNATQAE-----VEQALIKOVMAAKQALNGNAN----- 3120
QY 195 LKRHDKVDDLSKVGLSNQCETLAKTIDNLNOAY 226
Db 3121 -VOHAKDEATALINSSVDLNOQAK-DALKQOV 3150

Search completed: March 6, 2006, 20:44:31
Job time : 164 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:22:57 ; Search time 230 Seconds
(without alignments)
699.393 Million cell updates/sec

Title: US-09-360-685C-25

Perfect score: 1127

Sequence: 1 KNGKNDFSKVTQAKSDLEN.....LSRNOELAQKIDNMQAVSE 228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	100.0	1147	1	CGA1_HELPY
2	1127	100.0	1147	1	P80200 helicobacte
3	1035	91.8	1198	2	O6VFM7_HELPY
4	1029	91.3	1181	2	O6VFM7_HELPY
5	1029	91.3	1181	2	O6VFM7_HELPY
6	1025.5	91.0	267	2	O4PL15_HELPY
7	1016	90.2	1142	2	O8RNU0_HELPY
8	1009	89.5	1180	2	O52GZ6_HELPY
9	1004	89.1	326	2	O8KZB0_HELPY
10	1003.5	89.0	1179	2	O7X4J1_HELPY
11	992	88.0	1186	2	O915X9_HELPY
12	990.5	87.9	1179	2	O52GZ5_HELPY
13	984	87.3	1182	2	O6VFM7_HELPY
14	982.5	87.2	394	2	O8KZB0_HELPY
15	972.5	86.3	327	2	O8KZB0_HELPY
16	972.5	86.3	327	2	O8KZB0_HELPY
17	963	85.4	441	2	O917K3_HELPY
18	962.5	85.4	359	2	O8KZB8_HELPY
19	962.5	85.4	359	2	O8KZB8_HELPY
20	960	85.2	1183	2	O60F05_HELPY
21	959.5	85.1	1186	1	O8KZB4_HELPY
22	958.5	85.0	1147	1	CAGA_HELPY
23	958.5	85.0	1147	1	O9F223_HELPY
24	955.5	84.8	1156	2	O9F115_HELPY
25	952	84.5	1216	2	O5D6E2_HELPY
26	948.5	84.2	1247	2	O8KZB9_HELPY
27	944.5	83.8	408	2	O917K4_HELPY
28	942.5	83.6	1184	2	O60F02_HELPY
29	941	83.5	1222	2	O8KZB8_HELPY
30	939.5	83.4	359	2	O8KZB4_HELPY
31	939.5	83.4	1183	2	O60F05_HELPY

32	939.5	83.4	1183	2	O60FP7_HELPY	O60FP7 helicobacte
33	939.5	83.4	1183	2	O60FP3_HELPY	O60FP3 helicobacte
34	939.5	83.4	1183	2	O60FP3_HELPY	O60FP3 helicobacte
35	938.5	83.3	1190	2	O5D6E1_HELPY	O5D6E1 helicobacte
36	937.5	83.2	1183	2	O60FP9_HELPY	O60FP9 helicobacte
37	934	82.9	391	2	O8KZC1_HELPY	O8KZC1 helicobacte
38	934	82.9	1214	2	O60F08_HELPY	O60F08 helicobacte
39	930.5	82.6	362	2	O8KZB5_HELPY	O8KZB5 helicobacte
40	930.5	82.6	1183	2	O60F04_HELPY	O60F04 helicobacte
41	930.5	82.6	1189	2	O60F04_HELPY	O60F04 helicobacte
42	926.5	82.2	1320	2	O6VFM7_HELPY	O6VFM7 helicobacte
43	918	81.5	1184	2	O60FP2_HELPY	O60FP2 helicobacte
44	912.5	81.0	445	2	O917K5_HELPY	O917K5 helicobacte
45	902.5	80.1	1182	1	CAGA_HELPY	P35746 helicobacte

ALIGNMENTS

RESULT 1
CGA1_HELPY
ID CGA1_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN Name=caga; Synonyms=cag;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N., Rappunli R.,
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal ulcer."
RT Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108; DOI=10.1073/pnas.93.25.14648;
RA Censini S., Lange C., Xiang Z., Crabtree J., Chiara P., Borodovsky M.,
RA Rappunli R., Covacci A.;
RT "cag, a pathogenicity island of Helicobacter pylori, encodes type I-
RT specific and disease-associated virulence factors."
RT Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RN [3]
RP PROTEIN SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506;
RP 661-677; 900-914 AND 1062-1077.
RA Herrmann V., Herrmann U., Kist M.;
RT Submitted (APR-1993) to Swiss-Prot.
RT -1- FUNCTION: May be necessary for the transcription, folding, export,
RT or function of the cytotoxin.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL: X70039; CAA49633.1; -; Genomic DNA.
EMBL: AF282853; AAC44706.1; -; Genomic DNA.
PIR: B48281; B48281.
InterPro: IPR005163; Caga.
InterPro: IPR004355; IVSec_caga.
PIfam: PF03507; Caga; 1.
DR PRINTS: PRO1553; TYB48SCAGA.
KW Antigen; Direct protein sequencing.
KW COMBIDAS 880 885 Poly-Asn.

FT CONFLICT 320 320 G -> A (1n Ref. 3).
FT CONFLICT 325 325 P -> F (1n Ref. 3).
FT CONFLICT 328 328 R -> K (1n Ref. 3).
FT CONFLICT 426 426 K -> E (1n Ref. 3).
FT CONFLICT 429 429 O -> E (1n Ref. 3).
FT CONFLICT 673 675 A -> T (1n Ref. 3).
FT CONFLICT 901 901 A -> T (1n Ref. 3).
FT CONFLICT 903 903 L -> R (1n Ref. 3).
FT CONFLICT 907 907 L -> P (1n Ref. 3).
FT CONFLICT 910 910 P -> R (1n Ref. 3).
FT CONFLICT 914 914 P -> E (1n Ref. 3).
FT CONFLICT 1072 1072 P -> S (1n Ref. 3).
FT CONFLICT 1074 1074 S -> D (1n Ref. 3).
SQ SEQUENCE 1147 AA; 128014 MW; AB92770835F68490 CRC64;

Query Match 100.0%; Score 1127; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 2, 1e-58;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 809
DB 61 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 810 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 869
DB 121 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 870 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 929
DB 181 ASGLGVGOAGGFLPKRHDKVDLSKVGLSRNQELAKIDRLNOI 929
DB 930 ASGLGVGOAGGFLPKRHDKVDLSKVGLSRNQELAKIDRLNOI 977

RESULT 2
ID 09F220 HELPY PRELIMINARY; PRT; 1147 AA.
AC 09F220
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Caga.
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
RN [1]_TaxID=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NC11637;
RX MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s00550070002;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchiyama H.,
Ando T.,
RT "heterogeneity found in the caga gene of Helicobacter pylori from
Japanese and non-Japanese isolates";
RI J. Gastroenterol. 35:890-897(2000).
DB BMBJ; AB015416; BAB20926.1; -? Genomic DNA.
DB GO; GO:0019534; F:toxin transporter activity; IEA.
DB InterPro; IPR005169; Caga.
DB InterPro; IPR004355; IVSec_caga.
DB Pfam; PF03507; Caga; 1.
DB PRINTS; PRO1553; TYPE4SCAGA.
SQ SEQUENCE 1147 AA; 128003 MW; D740A727549B141D CRC64;

Query Match 100.0%; Score 1127; DB 2; Length 1147;
Best Local Similarity 100.0%; Pred. No. 2, 1e-58;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 809

DB 61 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 810 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 869
DB 121 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 870 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 929
DB 181 ASGLGVGOAGGFLPKRHDKVDLSKVGLSRNQELAKIDRLNOI 929
DB 930 ASGLGVGOAGGFLPKRHDKVDLSKVGLSRNQELAKIDRLNOI 977

RESULT 3
ID 06VROS HELPY PRELIMINARY; PRT; 1198 AA.
AC 06VROS;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Caga.
GN Name=caga; ORFNames=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
RN [1]_TaxID=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CA2;
RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
RA Blombergren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.;
RT "Comparative analysis of the complete cag pathogenicity island
sequence in four Helicobacter pylori isolates";
RL Gene 328:85-93(2004).
DB EMBL; AT330637; AAR03881.1; -? Genomic DNA.
DB GO; GO:0019534; F:toxin transporter activity; IEA.
DB InterPro; IPR005169; Caga.
DB InterPro; IPR004355; IVSec_caga.
DB Pfam; PF03507; Caga; 1.
DB PRINTS; PRO1553; TYPE4SCAGA.
SQ SEQUENCE 1198 AA; 133091 MW; DBF5B554663BC0E CRC64;

Query Match 91.8%; Score 1035; DB 2; Length 1198;
Best Local Similarity 81.7%; Pred. No. 6, 2e-53;
Matches 214; Conservative 4; Mismatches 10; Indels 34; Gaps 1;

DB 1 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVVDYINOKTVDKVNINQAVSVAKATGDFSRVQALAD 825
DB 61 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 120
DB 826 LKNFSKEOLAQAQAKNESLNARKSEIYOSVKNVGVGLVNGLSQAETTLISKNFSDIK 885
DB 121 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 180
DB 886 KEINAKLGPNNNNNNGLNKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNOI 945
DB 181 ASGLGVGOAGG-----FPKRDHDKVDLSK 206
DB 946 ASGLGVGOAGGFLPKRHDKVDLSKVGLSRNQELAKIDRLNOI 977
DB 207 VGLSRNQLAKIDRLNOI 228
DB 1006 VGLSRNQLAKIDRLNOI 1027

RESULT 4
ID 06VRM7 HELPY PRELIMINARY; PRT; 1181 AA.
AC 06VRM7;
DT 05-JUL-2004 (TEMBLrel. 27, Created)

DT 05-JUL-2004 (Tremblrel. 27, last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, last annotation update)
 DE Caga.
 GN Name=caga; ORFNames=HP0547;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C473.
 RA Blomstedt=15019987; DOI=10.1016/j.gene.2003.11.029;
 RX Blomstedt A., Lundin A., Nilsson C., Engstrand L., Lundberg J.,
 RT "Comparative analysis of the complete cag pathogenicity island
 RT sequence in four Helicobacter pylori isolates."
 RL Gene 328:85-93(2004).
 DR EMBL: AY330639; AKR03909.1; -; Genomic_DNA.
 DR GO: GO:0019534; P:toxin transporter activity; IBA.
 DR InterPro: IPR005169; Caga.
 DR InterPro: IPR004355; IVSec_caga.
 DR Pfam: PF03507; Caga; 1.
 DR PRINTS: PR01553; TYPR45SCGA.
 SQ SEQUENCE 1181 AA; 131427 MW; A991CAF6D17CDE3B CRC64;

Query Match 91.3%; Score 1029; DB 2; Length 1181;
 Best Local Similarity 81.3%; Pred. No. 1.4e-52;
 Matches 213; Conservative 4; Mismatches 11; Indels 34; Gaps 1;

QY 1 KNGKNDPSKVTQAKSDLENSVVDVITINQKTDKVDNLNQAVSAKATGDFSRVEQALAD 60
 DB 750 KNGKNDPSKVTQAKSDLENSIKVDIINQKITDKVNLSAASVAKATGDFSRVEQVLNG 809
 QY 61 LKNFSKEQLAQQAOKNESINARKSEIYOSVKNVNGTLVNGLSQAERTTTSKNSDIX 120
 DB 810 LKNFSKEQLAQQAOKNESDFVTGKSEIYOSVKNVNGTLVNGLSQAERTTTSKNSDIX 869
 QY 121 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNLI 180
 DB 870 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNLI 929
 QY 181 ASGLGVVGOAAG-----FPLKRHDKVDLSK 206
 DB 930 ASGLGVVGOAAGFPLKRHDKVDLSKVGLSANPEPIYATIDELGDFPLKRHDKVDLSK 989
 QY 207 VGLSRNQLAQKIDNLNOAVSE 228
 DB 990 VGLSRNQLAQKIDNLNOAVSE 1011

RESULT 5

Q4PLIS_HELPY PRELIMINARY; PRT; 1230 AA.
 ID Q4PLIS_HELPY PRELIMINARY; PRT; 1230 AA.
 AC Q4PLIS;
 DT 13-SEP-2005 (Tremblrel. 31, Created)
 DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
 DE Cytoxin-associated protein A.
 GN Name=caga;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=7d9e;
 RA Kim S.Y., Blaaser M.J., Lee Y.C., Pillinger M.H.,
 RT "Helicobacter pylori stimulates matrix metalloproteinase-1 secretion
 RT from gastric epithelial cells via Caga-dependent and independent
 RT mechanisms: requirement for Etk activation."
 RL Submitt (May-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL: DD067454; AAY68033.1; -; Genomic DNA.
 SQ SEQUENCE 1230 AA; 136662 MW; 64B240B8E3001351 CRC64;

Query Match 91.3%; Score 1029; DB 2; Length 1230;
 Best Local Similarity 69.0%; Pred. No. 1.4e-52;
 Matches 218; Conservative 5; Mismatches 5; Indels 88; Gaps 1;

QY 1 KNGKNDPSKVTQAKSDLENSVVDVITINQKTDKVDNLNQAVSAKATGDFSRVEQALAD 60
 DB 746 KNGKNDPSKVTQAKSDLENSIKVDIINQKITDKVNLSAASVAKATGDFSRVEQALAD 805
 QY 61 LKNFSKEQLAQQAOKNESINARKSEIYOSVKNVNGTLVNGLSQAERTTTSKNSDIX 120
 DB 806 LKNFSKEQLAQQAOKNESFVVGKSEIYOSVKNVNGTLVNGLSQAERTTTSKNSDIX 865
 QY 121 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNLI 180
 DB 866 KEINAKLGPNFNNNNNGLNKEPIYAKVKKKGAQAASLEPIYAQVAKKVNKIDRLNLI 925
 QY 181 ASGLGVVGOAAG-----FPLKRHDKVDLSK 192
 DB 926 ASGLGVVGOAAGFPLKRHDKVDLSKVGSRVSEPIYATIDELGDFPLKRHDKVDLSK 985
 QY 193 -----FPLKRHDKVDLSK 212
 DB 986 VGOAAGFPLKRHDKVDLSKVGSRVSEPIYATIDELGDFPLKRHDKVDLSK 1045
 QY 213 QELAOKIDNLNOAVSE 228
 DB 1046 QELAOKIDNLNOAVSE 1061

RESULT 6

Q8RNDU_HELPY PRELIMINARY; PRT; 267 AA.
 ID Q8RNDU_HELPY PRELIMINARY; PRT; 267 AA.
 AC Q8RNDU;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Cytoxin-associated antigen (Fragment).
 GN Name=caga;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 NCBI_TaxID=210;
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=CH-CTX1;
 RX MEDLINE=20123532; PubMed=10660136;
 RA Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
 RT "Serological response to Helicobacter pylori recombinant antigens in
 RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
 RT gastric cancer."
 RL APMIS 107:1069-1078 (1999).
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=CH-CTX1;
 RX MEDLINE=22121498; PubMed=12125208;
 RA Bruce R., Mancilla M., Valenzuela P., Yudelevich A., Vengas A.,
 RT "Cloning and comparison of ten gene sequences of a Chilean H. pylori
 RT strain with other H. pylori strains revealed higher variability for
 RT vacA and caga virulence factors."
 RL Biol. Res. 35:67-84 (2002).
 DR EMBL: AF479032; AAL6902.1; -; Genomic_DNA.
 DR InterPro: IPR005169; Caga.
 DR Pfam: PF03507; Caga; 1.
 FT NON_TER 1
 FT NON_TER 267
 SQ SEQUENCE 267 AA; 29032 MW; 04B4F51480960E8B CRC64;

Query Match 91.0%; Score 1025.5; DB 2; Length 267;
 Best Local Similarity 81.5%; Pred. No. 4.5e-53;
 Matches 216; Conservative 5; Mismatches 7; Indels 37; Gaps 3;

```

QY 1 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 3 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 62
QY 61 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 120
DB 63 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 122
QY 121 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRL 177
DB 123 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRL 182
QY 178 NOIASGLGVGGAAG-----PILKRHDKYDD 203
DB 183 NOIASGLGVGGAAGFPLKRHDKYDDLSKVGSRVSPPIYATIDDLGGFPLKRHDKYDD 242
QY 204 LSKVGLSHNOELAQKIDNLNOAVSE 228
DB 243 LSKVGLSHNOELAQKIDNLNOAVSE 267

```

```

RESULT 7
ID 052GZ6_HELPY PRELIMINARY; PRT; 1142 AA.
AC 052GZ6_HELPY
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Caga (Fragment)
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2808.
RA PubMed=15972330; DOI=10.1073/pnas.0409873102;
RT "NF- $\kappa$ B activation and potentiation of proinflammatory responses
RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
DR EMBL; D0011619; AY18597.1; Genomic DNA.
DR GO; GO:0019534; F10xin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
FT NON TER 1142
SQ SEQUENCE 1142 AA; 128233 MW; E53DBE9756EA29BE CRC64;

```

```

Query Match 90.2%; Score 1016; DB 2; Length 1142;
Best Local Similarity 89.0%; Pred No. 7, 8e-52;
Matches 203; Conservative 7; Mismatches 18; Indels 0; Gaps 0;
QY 1 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 745 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 804
QY 61 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 120
DB 805 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 864
QY 121 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 180
DB 865 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 924
QY 181 ASGLGVGGAAGFPLKRHDKYDDLSKVGSRVSPPIYATIDDLGGFPLKRHDKYDDLSK 228
DB 925 ASGLGVGGAAGFPLKRHDKYDDLSKVGSRVSPPIYATIDDLGGFPLKRHDKYDDLSK 972

```

RESULT 8
08KZK3_HELPY

```

ID 08KZK3_HELPY PRELIMINARY; PRT; 1180 AA.
AC 08KZK3_HELPY
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE Caga (Cag pathogenicity island protein).
CN Name=Caga; Synonyms=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112;
RA MEDLINE=2235025; PubMed=12391297; DOI=10.1073/pnas.222375399;
RA Higaehi H., Teutsumi R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
RA Harakeyama W.,
RT "Biological activity of the Helicobacter pylori virulence factor Caga
is determined by variation in the tyrosine phosphorylation sites."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112.
RA Yamakawa A., Yamazaki S., Azuma T.,
RA Yamakawa A.,
RN Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112;
RA Azuma T., Yamakawa A., Yamazaki S.,
RL Submitted (SRP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB090088; BAC10434.1; Genomic DNA.
DR EMBL; AB120425; BAC10434.1; Genomic DNA.
DR GO; GO:0019534; F10xin transporter activity; IEA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1180 AA; 130973 MW; 623474AF2635020 CRC64;

```

```

Query Match 89.5%; Score 1009; DB 2; Length 1180;
Best Local Similarity 79.0%; Pred No. 2, 1e-51;
Matches 207; Conservative 9; Mismatches 12; Indels 34; Gaps 1;
QY 1 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVGVYINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 809
QY 61 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 120
DB 810 LKNFSKEQLAQQAOKNESLNARKKSEIYQSVKNGVNGTLVNGLSQAEATTLTKNFSFDIK 869
QY 121 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 180
DB 870 KEINAKLGNFNNNNNNNGLNKNPIYAKVKKKGAQAASLEPIYAOVAKKNAKIDRLNOI 929
QY 181 ASGLGVGGAAG-----PILKRHDKYDDLSK 206
DB 930 ASGLGVGGAAGFPLKRHDKYDDLSKVGSRVSPPIYATIDDLGGFPLKRHDKYDDLSK 969
QY 207 VELSHNOELAQKIDNLNOAVSE 228
DB 990 VELSHNOELAQKIDNLNOAVSE 1011

```

```

RESULT 9
ID 07XKJ1_HELPY PRELIMINARY; PRT; 326 AA.
AC 07XKJ1_HELPY
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE Caga (Fragment).
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK112.
RA Yamakawa A., Yamazaki S., Azuma T.,
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

```

OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=China:079a;
 RX PubMed=14734191; DOI=10.1016/S0928-8244(03)00299-2;
 RA Zhou W., Yamazaki S., Yamakawa A., Ohtani M., Ico Y., Keida Y.,
 Higaishi H., Hatakeyama M., Si J., Azuma T., Ico Y., Keida Y.,
 RT "The diversity of vacA and cagA genes of Helicobacter pylori in East
 Asia".
 RL FEMS Immunol. Med. Microbiol. 40:81-87 (2004).
 DR EMBL; AB110963; BAC77006.1; -; Genomic_DNA.
 DR InterPro; IPR005169; Caga.
 DR Pfam; PF03507; Caga; 1.
 FT NON_TER 1 326
 FT NON_TER 326 326
 SQ SEQUENCE 326 AA; 35988 MW; 7BC601BFA98ACABF CRC64;

Query Match 89.1%; Score 1004; DB 2; Length 326;
 Best Local Similarity 88.2%; Pred. No. 1e-51;
 Matches 201; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 60
 DB 57 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 116
 QY 61 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 117 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 176
 QY 121 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 180
 DB 177 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 236
 QY 181 ASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 228
 DB 237 ASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 284

RESULT 10
 Q9LSX9 HELPY PRELIMINARY; PRT; 1179 AA.
 AC Q9LSX9
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Cytotoxin-associated protein CagA.
 GN Name=cagA;
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SS1;
 RA Zhang J.Z.;
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=SS1;
 RA Zhou J.C., Zhang J.Z.;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF244651; AAF63759.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; Caga.
 DR InterPro; IPR004355; IVSec_caga.
 DR Pfam; PF03507; Caga; 2.
 DR PRINTS; PRO1553; TYPE4SSCAGA.
 SQ SEQUENCE 1179 AA; 131297 MW; 7A9112F58B749787 CRC64;

Query Match 89.0%; Score 1003.5; DB 2; Length 1179;

Best Local Similarity 79.8%; Pred. No. 4.4e-51;
 Matches 209; Conservative 5; Mismatches 13; Indels 35; Gaps 2;
 QY 1 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 60
 DB 750 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 809
 QY 61 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 810 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 869
 QY 121 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 180
 DB 870 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 929
 QY 181 ASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 228
 DB 930 ASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 988
 QY 207 VGLSRNOELAKQKIDNLNOAVSE 228
 DB 989 VGLSRNOELAKQKIDNLNOAVSE 1010

RESULT 11
 Q52GZ5 HELPY PRELIMINARY; PRT; 1186 AA.
 AC Q52GZ5
 DT 13-SEP-2005 (Tremblrel. 31, Created)
 DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
 DE Caga (Fragment).
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxId=210;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=P310;
 RX PubMed=15972330; DOI=10.1073/pnas.0409873102;
 RA Brandt S., Kwok T., Hartig R., Konig W., Backert S.;
 RT "NF- κ B activation and potentiation of proinflammatory responses
 by the Helicobacter pylori Caga protein".
 RL Proc. Natl. Acad. Sci. U.S.A. 102:9300-9305 (2005).
 DR EMBL; DQ011620; AAY18598.1; -; Genomic DNA.
 DR GO; GO:0019534; F:toxin transporter activity; IEA.
 DR InterPro; IPR005169; Caga.
 DR InterPro; IPR004355; IVSec_caga.
 DR Pfam; PF03507; Caga; 1.
 DR PRINTS; PRO1553; TYPE4SSCAGA.
 FT NON_TER 1186 1186
 SQ SEQUENCE 1186 AA; 131865 MW; CDSAC9D1DA1BF3F CRC64;

Query Match 88.0%; Score 992; DB 2; Length 1186;
 Best Local Similarity 78.8%; Pred. No. 2.1e-50;
 Matches 212; Conservative 5; Mismatches 10; Indels 42; Gaps 3;

QY 1 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 60
 DB 750 KNGKNDPSKYTQAKSDLENSVDVIINOKYTDKVDNLNOAVSAKATGDFSRVEQALAD 809
 QY 61 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 120
 DB 810 LKNFSKEQLAQAQKNSLNARKKSEIYQSVKGVNGTLVNGLSQAETTLTKNFSIDIK 869
 QY 121 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 180
 DB 870 KEINAKLGNNFNNNNNGKLNKPEPIYAKVKKAGQAASLEPIYAKVKKAKIDRLNOI 929
 QY 174 IDRLNOIASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 228
 DB 929 IDRLNOIASGLGVGQAAGFPLKRHDKVDLSKVGSRNOELAKQKIDNLNOAVSE 988

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rup

Page 6

```
QY 200 KVDLSKYGLSRNOELAKINDLNQAVSE 228
DB 989 KVDLSKYGLSRNOELAKINDLNQAVSE 1017

RESULT 12
Q9F222_HELPY PRELIMINARY; PRT; 1179 AA.
AC Q9F222;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;

NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC43579;
RX MEDLINE=21457648; PubMed=11573724; DOI=10.1007/s005350070002;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchiyama H.,
RT "Heterogeneity found in the caga gene of Helicobacter pylori from
RL Japanese and non-Japanese isolates."
DR EMBL; AB015414; BAE20924.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga_1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1179 AA; 13168 MW; 7D0AC34E6446434 CRC64;

Query Match
Best Local Similarity 87.9%; Score 990.5; DB 2; Length 1179;
Matches 206; Conservative 5; Mismatches 17; Indels 33; Gaps 1;

QY 1 KNGKNDPSKYTOAKSDLENSVVDVYINQKVTDKVDNLNQAIVAKATGDSFREVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVVDVYINQKVTDKVDNLNQAIVAKATGDSFREVQALAD 809
QY 61 LKNFSKEQLAQAOKNESLNAKRSSEIYOSVKNGVNGTLVNGLSQAEATTLSSKNFSDIK 120
DB 810 LKNFSKEQLAQAOKNESFVNGKSEIYOSVKNGVNGTLVNGLSQAEATTLSSKNFSDIK 869
QY 121 KEINAKLGNFNNNNNGNLKNEPIYAKVKKKAGQAASLEPIYAOVAKVNAKIDRLN 180
DB 870 KEINAKLGNFNNNNNGNLKNEPIYAKVKKKAGQAASLEPIYAOVAKVNAKIDRLN 929
QY 181 ASGLGVVQQA-----AGFPLKRRHKKVDLSKY 207
DB 930 ASGLGVVQQAAGFPLKRRHKKVDLSKYGLSRNOELAKINDLNQAVSE 228
QY 208 GLSRNOELAKINDLNQAVSE 228
DB 990 GLSRNOELAKINDLNQAVSE 1010

RESULT 13
Q6VRG7_HELPY PRELIMINARY; PRT; 1182 AA.
AC Q6VRG7;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
GN Name=caga; ORFNames=HP0547;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;
```

```
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DJ23:2;
RX PubMed=15019987; DOI=10.1016/j.gene.2003.11.029;
RA Blomstergren A., Lundin A., Nilsson C., Engstrand L., Lundberg J.,
RT "Comparative analysis of the complete cag pathogenicity island
RL sequence in four Helicobacter pylori isolates."
DR EMBL; AY330644; AAR03970.1; -; Genomic DNA.
DR GO; GO:0019534; F:toxin transporter activity; IEA.
DR InterPro; IPR004355; Caga.
DR Pfam; PF03507; Caga_1.
DR PRINTS; PRO1553; TYPE4SSCAGA.
SQ SEQUENCE 1182 AA; 131615 MW; BDECS36BE4AF1E27 CRC64;
```

```
Query Match
Best Local Similarity 87.3%; Score 984; DB 2; Length 1182;
Matches 207; Conservative 5; Mismatches 16; Indels 36; Gaps 2;

QY 1 KNGKNDPSKYTOAKSDLENSVVDVYINQKVTDKVDNLNQAIVAKATGDSFREVQALAD 60
DB 750 KNGKNDPSKYTOAKSDLENSVVDVYINQKVTDKVDNLNQAIVAKATGDSFREVQALAD 809
QY 61 LKNFSKEQLAQAOKNESLNAKRSSEIYOSVKNGVNGTLVNGLSQAEATTLSSKNFSDIK 120
DB 810 LKNFSKEQLAQAOKNESFVNGKSEIYOSVKNGVNGTLVNGLSQAEATTLSSKNFSDIK 869
QY 121 KEINAKLGNFNNNNNGNLKNEPIYAKVKKKAGQAASLEPIYAOVAKVNAKIDRLN 178
DB 870 KEINAKLGNFNNNNNGNLKNEPIYAKVKKKAGQAASLEPIYAOVAKVNAKIDRLN 929
QY 179 QIASGLGVVQQAAGFPLKRRH-----KVDL 204
DB 930 QIASGLGVVQQAAGFPLKRRHKKVDLSKYGLSRNOELAKINDLNQAVSE 228
QY 205 SKVGLSRNOELAKINDLNQAVSE 228
DB 990 SKVGLSRNOELAKINDLNQAVSE 1013
```

```
RESULT 14
Q8RRY0_HELPY PRELIMINARY; PRT; 394 AA.
AC Q8RRY0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX NCBI_Taxid=210;

NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC51407;
RA Yamaoka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB057003; BBA81770.1; -; Genomic DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga_1.
DR NON TER 1.
FT NON TER 394
SQ SEQUENCE 394 AA; 43035 MW; BIC9D5596A8B6E5F CRC64;
```

```
Query Match
Best Local Similarity 87.2%; Score 982.5; DB 2; Length 394;
Matches 211; Conservative 5; Mismatches 10; Indels 55; Gaps 3;

1 KNGKNDPSKYTOAKSDLENSVVDVYINQKVTDKVDNLNQAIVAKATGDSFREVQALAD 60
```


Search completed: March 6, 2006, 20:29:47
Job time : 231 secs

```
Db 69 KNGKNDPSKYTOAKSDLENSVXDVIINOKITDKVDNLNOBVLVAKATGDFSWVEQALAD 128
QY 61 LKNFSKEQLAQOAKQKESLNARKKSEIYOSVKGNGVGTLYGNGLSQAEATTLTKNPSDIK 120
Db 129 LKNFSKEQLAQOAKQKESLNARKKSEIYOSVKGNGVGTLYGNGLSQAEATTLTKNPSDIK 188
QY 121 KEINAKLGNFNNNNNGLNK-----EPIYAKVKKKAGQAASLEEP 161
Db 189 KEINAKLGNFNNNNR--LKNEPIYAKVKKKAGQAASPEEPIYAKVKKKAGQAASPEEP 246
QY 162 IYAQVAKKVNAKIDRLNQIASGLGVGQAAGFPLKRHDKYDILS----- 205
Db 247 IYAQVAKKVNAKIDRLNQIASGLGVGQAAGFPLKRHDKYDILSKVGLSARSEPIYATID 306
QY 206 -----KVGLSRNOELAKIDNLNOAVSE 228
Db 307 GLGGPPLKRHDYDELIRKVLGSLRSTQKLAQKIDNLNOAVSE 347
```

```
RESULT 15
Q8KZB0_HELPY PRELIMINARY; PRT; 327 AA.
AC Q8KZB0;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Caga (Fragment).
GN Name=caga;
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RX MEDLINE=22295025; PubMed=12391297; DOI=10.1073/pnas.222375399;
RA Higaehi H., Teutsuni R., Fujita A., Yamazaki S., Asaka M., Azuma T.,
RA Hatakeyama M.;
RT "Biological activity of the Helicobacter pylori virulence factor Caga
is determined by variation in the tyrosine phosphorylation sites.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14428-14433(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OK181;
RA Yamakawa A., Yamazaki S., Azuma T.;
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB090151; BAC10497.1; -; Genomic_DNA.
DR InterPro; IPR005169; Caga.
DR Pfam; PF03507; Caga; 1.
FT NON_TER 1
FT NON_TER 327
SQ SEQUENCE 327 AA; 35798 MW; 1309C5E2418C7DEC CRC64;
```

Query Match 86.3%; Score 972.5; DB 2; Length 327;
Best Local Similarity 86.4%; Pred. No. 7.6e-50;
Matches 197; Conservative 10; Mismatches 20; Indels 1; Gaps 1;

```
QY 1 KNGKNDPSKYTOAKSDLENSVXDVIINOKITDKVDNLNOAVSVAKATGDFSRVEQALAD 60
Db 59 KNGKNDPSKYTOAKSDLENSVXDVIINOKITDKVDNLNOAVSIANSTDFSGVEQALAD 118
QY 61 LKNFSKEQLAQOAKQKESLNARKKSEIYOSVKGNGVGTLYGNGLSQAEATTLTKNPSDIK 120
Db 119 LKNFSKEQLAQOAKQKEDFNLTGNSSELYOSVKGNGVGTLYGNGLSQAEATTLTKNPSDIK 178
QY 121 KEINAKLGNFNNNNNGLNKNEPIYAKVKKKAGQAASLEEPIYAOVAKKVNAKIDRLNOI 180
Db 179 KEINAKLGNFNNNNNGLNKNEPIYAKVKKKAGQAASLEEPIYAOVAKKVNAKIDRLDOI 238
QY 181 ASGLGVGQAAGFPLKRHDKYDILSKVGLSRNOELAKIDNLNOAVSE 228
Db 239 ASGLGVGQAAGFPLKRHDKYDILSKVGLSRNOELAKIDNLNOAVSE 285
```

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2006, 20:41:53 ; Search time 21 Seconds

(without alignments)
217.147 Million cell updates/sec

Title: US-09-360-685C-25

Perfect score: 1127

Sequence: 1 KNGKNDPFSKYTQAKSDLEN.....LSRNOELAQKIDNINQAVSE 228

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 135346

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1127	100.0	1147	US-10-615-668-5	Sequence 5, Appli
2	896.5	79.5	1167	US-11-052-554A-121	Sequence 121, App
3	127	11.3	1290	US-10-485-517-141	Sequence 141, App
4	125	11.1	1448	US-10-485-517-212	Sequence 212, App
5	120.5	10.7	5024	US-10-793-626-2964	Sequence 2964, Ap
6	118	10.5	1713	US-10-766-317-2	Sequence 2, Appli
7	118	10.5	1724	US-10-766-317-6	Sequence 6, Appli
8	118	10.5	1724	US-10-766-317-4	Sequence 4, Appli
9	117	10.4	693	US-11-196-475-68	Sequence 68, Appli
10	115.5	10.2	1095	US-10-793-626-3154	Sequence 3154, Ap
11	114	10.1	3712	US-11-019-711-48	Sequence 48, Appli
12	112	9.9	3712	US-11-019-711-51	Sequence 51, Appli
13	111.5	9.9	700	US-11-196-475-66	Sequence 66, Appli
14	110.5	9.8	708	US-11-196-475-76	Sequence 76, Appli
15	110.5	9.8	739	US-11-087-099-12273	Sequence 12273, A
16	109.5	9.7	2087	US-11-075-185-28	Sequence 28, Appli
17	109	9.7	611	US-10-793-626-2586	Sequence 2586, Ap
18	109	9.7	2480	US-10-995-561-825	Sequence 825, App
19	109	9.7	3116	US-10-995-561-826	Sequence 826, App
20	108	9.6	401	US-11-097-749-3	Sequence 3, Appli
21	108	9.6	863	US-11-097-749-2	Sequence 74, Appli
22	105.5	9.4	700	US-11-196-475-74	Sequence 74, Appli
23	104.5	9.3	1404	US-10-878-556A-169	Sequence 169, App
24	104.5	9.3	1652	US-10-995-561-663	Sequence 663, App
25	104.5	9.3	1938	US-10-995-561-661	Sequence 661, App

26	104.5	9.3	1938	US-10-995-561-662	Sequence 662, App
27	104.5	9.3	1954	US-10-995-561-660	Sequence 660, App
28	104.5	9.3	1972	US-10-995-561-664	Sequence 664, App
29	104.5	9.3	1972	US-10-995-561-666	Sequence 666, App
30	103	9.1	440	US-10-976-933-2	Sequence 2, Appli
31	103	9.1	440	US-10-976-933-4	Sequence 4, Appli
32	102.5	9.1	1279	US-10-793-626-1188	Sequence 3188, Ap
33	102	9.1	663	US-11-196-475-70	Sequence 70, Appli
34	102	9.1	663	US-11-196-475-78	Sequence 78, Appli
35	101.5	9.0	380	US-11-144-833-13	Sequence 13, Appli
36	101	9.0	2036	US-11-124-368A-276	Sequence 276, App
37	101	9.0	2036	US-11-124-368A-281	Sequence 281, App
38	101	9.0	2036	US-11-124-368A-281	Sequence 281, App
39	101	9.0	2044	US-11-124-368A-278	Sequence 278, App
40	101	9.0	2147	US-10-995-561-827	Sequence 827, App
41	101	9.0	2147	US-11-124-368A-277	Sequence 277, App
42	100	8.9	794	US-10-793-626-1050	Sequence 1050, Ap
43	98.5	8.7	903	US-10-689-742-142	Sequence 142, App
44	98.5	8.7	1189	US-11-074-176-134	Sequence 134, App
45	98	8.7	239	US-10-485-517-249	Sequence 249, App

ALIGNMENTS

RESULT 1
US-10-615-668-5
Sequence 5, Application US/10615668
Publication No. US20050276819A1
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappelli, Rino
TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Polynucleotides
FILE REFERENCE: CHIR0337
CURRENT APPLICATION NUMBER: US/10/615,668
CURRENT FILING DATE: 2003-07-08
PRIOR APPLICATION NUMBER: 08/471,491
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/256,848
PRIOR FILING DATE: 1994-10-21
PRIOR APPLICATION NUMBER: 09/410,835
PRIOR FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-10-615-668-5

Query Match 100.0%; Score 1127; DB 6; Length 1147;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KNGKNDPFSKYTQAKSDLEN	VS	KVQVITKRVNLTNINQAVAKATGDFSRVEQALAD	60
DB	750	KNGKNDPFSKYTQAKSDLEN	VS	KNQVITKRVNLTNINQAVAKATGDFSRVEQALAD	809
QY	61	LKNSKQOLAQAQKNSLNARKKSEIYQSVKNGVNGTIVGNSLSQAEATTLTKNFSDIK			120
DB	810	LKNSKQOLAQAQKNSLNARKKSEIYQSVKNGVNGTIVGNSLSQAEATTLTKNFSDIK			869
QY	121	KEINAKLGNRNNNNNNGKLKKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNQT			180
DB	870	KEINAKLGNRNNNNNNGKLKKEPIYAKVKKKAGQAASLEPIYAQVAKVNAKIDRLNQT			929
QY	181	ASGIGVGAAGAPFLKRHDVDLTKVGLSRNOELAQKIDNINQAVSE 228			
DB	930	ASGIGVGAAGAPFLKRHDVDLTKVGLSRNOELAQKIDNINQAVSE 977			

US-10-763-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUA48005
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2964

Query Match 10.7%; Score 120.5; DB 6; Length 5024;
Best Local Similarity 23.2%; Pred. No. 4.7;
Matches 62; Conservative 50; Mismatches 84; Indels 71; Gaps 13;

QY 10 KVTQAKSLDLSVSKDVI-INCKVTDKVNLNOAVSAVATGDFSRVEQALDKKPSKQ 68
DB 74 KLTAKTETAINALNTLADLN--TPQKRAIKATITATRTDVT-AEOSKAKOINSAMHT 129
QY 69 LAQAQAKNES-----INARKK-----SEIYQSVKGVN--GTLVNGLSQ-AEATTL 112
DB 130 LRQNSIDNESVTNENYINAEPEKQHAFTALNNKKEIVNEQATLIDANSINQKQALIT 189
QY 113 SKNPSIDIKEL-----NA-----KLGNNNNNNNGLKNEPIYAKVKKKAGQASLE 159
DB 190 TKNALDGEOLRAKENADQEIINTLOLTDQORSEKGLVNS--SQRTTEVASQLAK-- 244
QY 160 EPIYQAVAKKNVAKIDRLNOIASGLGVGQAAGF-----PLKRRDK 200
DB 245 -----AKELNKVEQLNLNLINGKQMINSKFLINEDANOQAVSALASAEVLKNSQ 297
QY 201 VDDLSKVGLSRNOELAKQIDNLNOAVS 227
DB 298 NPELDKV-----TTEQAINNNINSAIN 318

RESULT 6
US-10-766-317-2
; Sequence 2, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 2
; LENGTH: 1713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-2

Query Match 10.5%; Score 118; DB 6; Length 1713;
Best Local Similarity 21.2%; Pred. No. 1.8;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

QY 5 NKDPSKYTOAKSLDLSVSKDVIINCKVTDKVNLNOAVSAVATGDFSR-----VEQALAD 60
DB 300 NNNVNRATQSAKELDVTKIKVIRNVHILKQISGTDEGNVVPDPSRMAEAQRMRE 359
QY 61 L--KNFSKEQLAQAQAKNESINARKKSEIYQSVKGVNGLVNGLSQAQEAATLSK--N 115
DB 360 LRNNFNGHLEAEADRGESQLLNRIRTWQKTHQGEN-----NGLANSIRDSINREYAK 414
QY 116 FSDIKKELNAGLGNFNNNNNNGLKNEPIYAKVKK-----KAGQASLE 160
DB 415 LSDLRARLOEAAQAQAKANGINQENRALGAIQHQVEINSLQSDFTKYTLTTADSSILQT 474
QY 161 PIYQAVAKKNVAKIDRLNOIASGLGVGQAAGFPLKRRDKVDLS-----KVGLSRN----- 212
DB 475 NIALQLEKSOKEYEKL--AASINLEARQ-----ELSDKVRLELSRAGKTSIVEAEK 525
QY 213 -----QELAKQIDNLNOAVS 227
DB 526 ARSLQELAKQLEIKRNAS 544

RESULT 7
US-10-766-317-6

; Sequence 6, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 6
; LENGTH: 1724
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-6

Query Match 10.5%; Score 118; DB 6; Length 1724;
Best Local Similarity 21.2%; Pred. No. 1.8;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

QY 5 NKDPSKYTOAKSLDLSVSKDVIINCKVTDKVNLNOAVSAVATGDFSR-----VEQALAD 60
DB 311 NNNVNRATQSAKELDVTKIKVIRNVHILKQISGTDEGNVVPDPSRMAEAQRMRE 370
QY 61 L--KNFSKEQLAQAQAKNESINARKKSEIYQSVKGVNGLVNGLSQAQEAATLSK--N 115
DB 371 LRNNFNGHLEAEADRGESQLLNRIRTWQKTHQGEN-----NGLANSIRDSINREYAK 425
QY 116 FSDIKKELNAGLGNFNNNNNNGLKNEPIYAKVKK-----KAGQASLE 160
DB 426 LSDLRARLOEAAQAQAKANGINQENRALGAIQHQVEINSLQSDFTKYTLTTADSSILQT 485
QY 161 PIYQAVAKKNVAKIDRLNOIASGLGVGQAAGFPLKRRDKVDLS-----KVGLSRN----- 212
DB 486 NIALQLEKSOKEYEKL--AASINLEARQ-----ELSDKVRLELSRAGKTSIVEAEK 536
QY 213 -----QELAKQIDNLNOAVS 227
DB 537 ARSLQELAKQLEIKRNAS 555

RESULT 8
US-10-766-317-4

; Sequence 4, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317

Query Match 10.5%; Score 118; DB 6; Length 1713;
Best Local Similarity 21.2%; Pred. No. 1.8;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

Tue Mar 7 12:58:24 2006

us-09-360-685c-25.rapbn

Page 4

```

/ CURRENT FILING DATE: 2004-01-27
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 4
/ LENGTH: 3333
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-766-317-4

```

```

Query Match
Best Local Similarity 10.5%; Score 118; DB 6; Length 3333;
Matches 55; Conservative 47; Mismatches 107; Indels 50; Gaps 9;

```

```

QY 5 NKDSKYTQAKSDLENSVDYIINQKVTDKVDNLNOAVSAKATGDSR-----VEQALAD 60
DB 1920 NNNVNRATQSAKEIDVYKIKVIRNVHILKQIGTGDBGNVPSGDSREMAAQRMRE 1979
QY 61 L--KNFSKQOLAQAQAKNESLNARKSEIYOSVKNVGNGLVNGLSQAEATLISK--N 115
DB 1980 LRRNRNGHLEAPADRESQULLNIRIRIQKTHQEN-----NGLNSIRDSINEYEAK 2034
QY 116 FSDIKKEINATLGNFNNNNNGLNKNEPIYAKVVK-----KAGQASLEB 160
DB 2035 LSDLRRLQEMAAQAQAQNGLNQENBERALGAIORQVEKINSIGSDPTKYLTTADSSILQT 2094
QY 161 PIYAQVAKVNAKIDNLNOASGLGVGAAGFPLKHDVDDLS-----KGLSRN--- 212
DB 2095 NIALQLWESQKYEKL-----AASLNBARQ-----ELSDKVELRSAGKTSIVEAEKH 2145
QY 213 -----OELAKIDNLNOAVS 227
DB 2146 ARSLQELAKQLEIRKNNAS 2164

```

```

RESULT 9
US-11-196-475-68
/ Sequence 68, Application US/11196475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Constructs of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ PRIOR FILING DATE: 2005-08-03
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR FILING DATE: 2001-08-07
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: PacSeq for Windows Version 4.0
/ SEQ ID NO 68
/ LENGTH: 693
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-68

```

```

Query Match
Best Local Similarity 10.4%; Score 117; DB 7; Length 693;
Matches 57; Conservative 53; Mismatches 91; Indels 62; Gaps 12;
QY 7 DESKYTQAKSDLENS--VDYIINQKVTDKVDNLNOAVSAKATGDS-----RVQO 56
DB 301 DLDRAQQLDSSSDNLDIQDVTYAEKIQIDIEIKENKLPK-PDVSVPVVDKQIQIE 359

```

```

QY 57 ALADLNKFSKQOLAQAQAKNESLNARKSEIYOSVKN-----GVNGTLVNGLSQ 106
DB 360 SLEDL-----QSLKETSIDENQREIEKQIEIKSDBELSKDPKALDNLGDL--NSKVS 413
QY 107 AEATYLSKNFSDIKKEINAKLGNFNNNNNGLNKNEPIYAKVNAKKAQAASLEPIYAOV 166
DB 414 SKKIKRGEIEVESKASLADLNEN-----LMEPEDQLSKDKLDSKQNLKPFV---- 466
QY 167 AKVNAKIDNLNOA-SGLGVGAAGFPLKHDVDDLSKVG----- 209
DB 467 -----SELEVNIEISKSNNEISS--PLYPSYSIDMSKEDIDNKDVLQETKSQTKS 519
QY 210 ---SRNQL-AQKIDNLNOAVSE 228
DB 520 QPSTLNQDLTTMSIDSSNPVPLE 542

```

```

RESULT 10
US-10-793-626-3154
/ Sequence 3154, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ PRIOR FILING DATE: 2004-03-04
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 3154
/ LENGTH: 1095
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3154

```

```

Query Match
Best Local Similarity 10.2%; Score 115.5; DB 6; Length 1095;
Matches 52; Conservative 33; Mismatches 78; Indels 53; Gaps 9;
QY 29 QKVTVDVNDLNOAV-----SVAKATGDSRVBQALADLKN-----FSKEQLAQ---QAQ 74
DB 14 OKAEQVIOVNAKKEIENAGQSVQVSDKSKVEQALSEINNAKSALADQKQLQAHVQLI 73
QY 75 KNSLNARKSEI-----YQSVKNVGNGLVNGLSQAEATLISKNFSDIKKEINAKKA 128
DB 74 QPTDNLNKKPASITANQRTQOFSNELNST-----KNTDRILKQNSVA 119
QY 129 NFNNNNNNGLNKNEPIYAKVNAKKAQAASLEPIYAOVAKVNAKIDNLNOASGLGVG 188
DB 120 DVNNALN---KRAEYQKLEAREALQNKEDSALVAKXQLQAVD---QVSTGMMQ 173
QY 189 QAAQFPLKHDVDDLSKVGSRNQLA-----OKIDN 221
DB 174 QTK-----DLYNSKQQAQAQOHSKQAQOVIDN 199

```

```

RESULT 11
US-11-019-711-48
/ Sequence 48, Application US/11019711
/ Publication No. US20060009634A1
/ GENERAL INFORMATION:
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Alabrook II, John P
/ APPLICANT: Tchernov, Vellizar T
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A
/ APPLICANT: Paturajan, Meera

```

APPLICANT: Grose, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malenker, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Bisen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/11/019,711
CURRENT FILING DATE: 2004-12-21
PRIOR APPLICATION NUMBER: US/10/037,417
PRIOR FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 48
LENGTH: 3712
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-11-019-711-48

Query Match 10.1%; Score 114; DB 7; Length 3712;
Best Local Similarity 20.7%; Pred. No. 8.9;
Matches 67; Conservative 46; Mismatches 100; Indels 110; Gaps 15;

5 NNDPSYVTAQSKLSESVQV---ITN-----OKVTDXNDINQAVS----- 43
2348 NSEKFTVSEQKQAEKNIDAGNGLTNGDLTLNQIKDLNLDALNEISFNKNVDEEL 2407
44 -----VAKATGDSRVEQALADLNKFSKQQLAQQAQKNBSLN-ARKKSELYGVKNG 94
2408 PVEDQKRAADLTID--QAEQKAAELAIKQDLAAQYTMVSAEPAITAAATVSGIVEA 2465
95 VNG-----TLVNG-----LSQAEATTLSKNSFSDIKKEINA 125
2466 VEAAGKLSQDAISAAGNATDKTDGIBRAHLADYTGSTDLLQARQSLQVQDDIEPRANA 2525
126 -----KLGNNNNNNNGKNEPIYAKVNF-----KKAQAASLE--- 159
2526 SAGKYOKISAVNNATHEHQLKD-----INKLIDQLPAESQDRMWNKSNANASDALEILKN 2579
160 -----EPIYAOVAKK-----VNAKIDRLNG-IASGLGVVGAAGFPLRHHKVDLS-- 205

Db 2580 VAEILEPVSQVTKKELEKHAHCIRRDLDLTNKDVQANKQJDDVESSVKNLEIAEDIEEG 2639
QY 206 --RVGLSRNQELAQKIDINQAV 226
Db 2640 QHRVG--SQSRQLGQREIKAKQV 2661

RESULT 12
US-11-019-711-51
Sequence 51, Application US/11019711
Publication No. US2006009634A1
GENERAL INFORMATION:
APPLICANT: Kekuda, Ramesh
APPLICANT: Alsobrook II, John P
APPLICANT: Tchernev, Velizar T
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A
APPLICANT: Baturajan, Meera
APPLICANT: Grose, William M
APPLICANT: Lepley, Denise M
APPLICANT: Burgess, Catherine E
APPLICANT: Vernet, Corine A.M.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Edinger, Shlomit R
APPLICANT: Sciore, Paul
APPLICANT: Ellerman, Karen
APPLICANT: Malenker, Uriel M
APPLICANT: Rothenberg, Mark
APPLICANT: Stone, David J
APPLICANT: Boldog, Ferenc L
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh G
APPLICANT: Anderson, David W
APPLICANT: Padigaru, Muralidhara
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Miller, Charles E
APPLICANT: Bisen, Andrew J
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-235
CURRENT APPLICATION NUMBER: US/11/019,711
CURRENT FILING DATE: 2004-12-21
PRIOR APPLICATION NUMBER: US/10/037,417
PRIOR FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/260,018
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: 60/260,360
PRIOR FILING DATE: 2001-01-08
PRIOR APPLICATION NUMBER: 60/272,411
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: 60/272,817
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/291,186
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/303,231
PRIOR FILING DATE: 2001-07-05
PRIOR APPLICATION NUMBER: 60/305,060
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 60/318,405
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/318,700
PRIOR FILING DATE: 2001-09-12
NUMBER OF SEQ ID NOS: 227
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 51
LENGTH: 3712
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-11-019-711-51

Query Match 9.9%; Score 112; DB 7; Length 3712;
Best Local Similarity 21.1%; Pred. No. 12;

Tue Mar 7 12:58:24 2006

US-09-360-685c-25.rapbn

Page 6

Matches	68; Conservative	41; Mismatches	104; Indels	110; Gaps	14
Qy	5 NNDPFSKYLTAQKSPLENSYVYD	---11N---	QAYTRKYDNIINAQVS	-----	43
Db	2348 NKEPDTYSBOKIAOEKNIKDAGNFMFLINDGLTLYNQNDKIDNLRDLIELNSFRKNVSEEL	-----	-----	-----	43
Dy	44 -----	-----	-----	-----	240
Qy	2408 FREDQKHEKDALTY- QARQKAELATKAODLAQAYTDMTASAPRITAKATNSGVFA	-----	ARKKSEIYYVANG	94	
Qy	95 VNG-----	TLVNG-----	LSQAEATLTISKPSFEDIKETLN	125	
Db	2466 VEAQKLSQDLSAAGNATTKTDCIGEEBANLADYGTSTDLQRAFOISLOCPDDEPRLN	-----	-----	252	
Qy	126 -----	KLNPNNNNNKLNSEPIYAAKTK	-----	KKQGAASLE	159
Db	2526 SAGKVOKISAVNNATHEQLD-----	INCLIDLPRAESQDMKSNNSNNSADLTETLN	2579		
Qy	160 -----	EPPIYAQAQK-----	VNAKIDBLNQLASGLG-----	VYGQAQFPLKHHKRVDD	203
Db	2580 VLEETPEPVSUPTPELEKAGNIRNRLDTLTKVDSQANKQDLDDVESVSKLSELAIEDIEQ	-----	-----	2639	
Qy	204 LSAVQGISRQNDLQAKIDNTINQAV	226			
Db	2640 QHHVVS-SQSRQDQDELTENKQAV	2661			

```

RESULT 13
US-11-196-475-66
/ Sequence 66, Application US/11196475
/ Publication No. US20050271692A1
/ GENERAL INFORMATION:
/ APPLICANT: Datwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Constructs of Borrelia
/ FILE OF INVENTION: Burgdorferi
/ PRIORITY REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ PRIOR FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-15
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ PRIOR FILING DATE: 2001-08-07
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 66
/ LENGTH: 700
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-66

```

```

Query Match      9.99; Score 111.5; DB 7; Length 700;
Best Local Similarity 19.48; Freq. No. 1.67
Matches 50; Conservative 55; Mismatches 68; Indels 81; Gaps 10;

QY      5 NKOFSVYTOAKSDSENVKQVILNKOTVDKVDVNLQAVSVAKATGFS-----RVKQ 56
Db      303 DKAOQLTSLSDAINDL--VORNTREKIQEIDININENGLK- PGVSSPRVDKQLOIYE : : : :
QY      57 ALADLNKFSKEQLAQOQAQSLNARKKSSIIYSYVGNVQTLVNGVSGSQAEATVLLSKNF 116
Db      360 SLDEL--QQLKLETDENOKREIKREIQIETIKSDER-----LNSKDDKSKG 405
QY      117 S--DKKELANK-----LGNENNNNNNGKQIE-----PIYAKQ 149
          | : : | | | : |

```

Db 406 KALDREINLSAASAKSEKSAKEEITTKGSKOKSLGIDLINDENLAMPDQCKPEYKXLS 465

QY 150 KKAQAASLEEEITVQYAKVNAKIDRLNOLASLGVGVOAGFPLKHPDRVDLSKVL 209

Dd 466 KKEKPEVPS-----EKKDKT-----FKSNNVGELSPDK 496

QY 210 SPNOELAKTIDUNOAS 227

Dd 497 SSYKDIDSKETVAKVN 514

```

14 RESULT 14
15 / Sequence 76, Application US/11196475
16 / Publication No. US20050271682A1
17 / GENERAL INFORMATION:
18 / APPLICANT: Dattwyler, Raymond J.
19 / APPLICANT: Gomes Solecki, Marcia J. C.
20 / APPLICANT: Iuft, Benjamin J.
21 / APPLICANT: Dunn, John J.
22 / TITLES OF INVENTION: Recombinant Constructs of Borrelia
23 / TITLES OF INVENTION: Burgdorferi
24 / PRIORITY REFERENCE: 2631.1001-011
25 / CURRENT APPLICATION NUMBER: US/11/196,475
26 / CURRENT FILING DATE: 2005-08-03
27 / PRIOR APPLICATION NUMBER: US 08/148,191
28 / PRIOR FILING DATE: 1993-11-01
29 / PRIOR APPLICATION NUMBER: US 08/235,836
30 / PRIOR FILING DATE: 1994-04-29
31 / PRIOR APPLICATION NUMBER: US 09/666,017
32 / PRIOR FILING DATE: 2000-09-19
33 / PRIOR APPLICATION NUMBER: 03-19
34 / PRIOR FILING DATE: 2000-08-28, 60/226,484
35 / PRIOR APPLICATION NUMBER: PCT/US01/24736
36 / PRIOR FILING DATE: 2001-08-07
37 / NUMBER OF SEQ ID NOS: 213
38 / SOFTWARE: PaasSBO for Windows Version 4.0
39 / SEQ ID NO 76
40 / LENGTH: 708
41 / TYPE: PRT
42 / ORGANISM: Borrelia burgdorferi
43 / US-11-196-475-76

```

```

Query Match      9.8%; Score 110.5; DB 7; Length 708;
Bee Local Similarity 22.4%; Pred. No. 1.9;
Matches 51; Conservative 54; Mismatches 94; Indels 39; Gaps 9;

QY      5 NKPSRYVQASDLSKNSKYCVINQKVTQDKVNLNQAVSAKATGDF-----RYEQ 56
Db      303 DKAOQLDLSADNLD--VQRTVREKIQEDINEIKRKINLPK--PDVASSPVKQLOLKE 359
QY      57 ALADLNKNSKQKLAQAQAKNSLARK-----SFIYVANGVNGTAVNGLSQAE 109
Db      360 SLDEL-----QQLKMGEDNQRKEIKQEIKGDEBELSKDGKGSADVLEALDLEFLS 145
QY      110 TTLSKNFSDIKE-----LNAKLGPNNNNGKLNKEPIYAKVKKRQAQAASLEBEIY 163
Db      416 KASSEKSKSYVEEIEITKGRPSLGDINDKMLPFD-----QKPEPKIDKSLD 467
QY      164 AQVAKVNAKIDRLINQIA-SGLGVGAQAAGPFLKRGDKYDLISKGLSRNQELMAOKID 220
Db      468 GKKEKPEVSEVEKLDKRTSKSNVNEVQKLSPLDKRSIDLD--SKEEVDKNAINLOKID 523

RESULT 15
US-11-087-099-12273
Sequence 12273, Application US/11087099
General Information No. US20060041961A1
Application Information:
Title: Coat Abad, Mark S. et al.
Title: Coat Invention: Genes and Uses for Plant Improvement
File Reference: 38-21(53450)B EP
Current Application Number: US/11/087,099

```


! CURRENT FILING DATE: 2005-03-22
! NUMBER OF SEQ ID NOS: 12464
! SEQ ID NO 12273
! LENGTH: 739
! TYPE: PRT
! ORGANISM: Streptococcus pneumoniae
US-11-087-099-12273

Query Match 9.8%; Score 110.5; DB 7; Length 739;
Best Local Similarity 23.1%; Pred. No. 2;
Matches 61; Conservative 43; Mismatches 83; Indels 77; Gaps 10;

QY 1 KNGKNDPFSKYTOAK--SDLENSVKDVI-INOKVTDKVDNINQAVSVAKATGDFSR----- 53
DB 365 ENSKKEBELSKTAEELTAAFEQPKDTLKEKVAEAEKKVEEAKKAKADQKEEDRRNTP 424
QY 54 -----VEQALADLKNPSKEQLAQQAQKNEESLNARKKSEIYQSVKNGVNGTLVNGLS 105
DB 425 TNYTKTLELEIASDVKVKKAELELVKEEANSRNEKIKQAKEVES-----K 473
QY 106 QAEATTLKNFSDIKK-ELNAKIGNFNNNNNNGLNKPIYAKVKKKAGQA-ASLEPTIY 163
DB 474 KAEATRLKIKTDRKKAEEBAKR-----KAESEKKAABAKQKVDAREY 517
QY 164 AQYAK--KYNAKIDRLNOLIASGLGVVQAAGFPLKSHDKVD-----DL 204
DB 518 ALEAKIAELEVEYQRIEK-----ELKEIDESDSSEDIKKEGLRAPLQSKLDT 563
QY 205 SKVGLSRNOELAQKIDNINQAVSE 228
DB 564 KKAKLSKLEBELSDKIDELDAEIAK 587

Search completed: March 6, 2006, 20:44:57
Job time : 21 secs

This Page Blank (uspto)